# Empirical likelihood test in a posteriori change-point nonlinear model

Gabriela Ciuperca, Zahraa Salloum

**Noname manuscript No.**
(will be inserted by the editor)

# Empirical likelihood test in a posteriori change-point nonlinear model

**Gabriela Ciuperca · Zahraa Salloum**

Received: date / Accepted: date

**Abstract** In this paper, in order to test whether changes have occurred in a nonlinear parametric regression, we propose a nonparametric method based on the empirical likelihood. Firstly, we test the null hypothesis of no-change against the alternative of one change in the regression parameters. The asymptotic behaviour of the empirical likelihood statistic under the null hypothesis and its alternative is studied. Under null hypothesis, the consistency and the convergence rate of the regression parameter estimators are proved. The critical value is chosen so that the test has a small probability of a false alarm and asymptotic power one. The epidemic model, a particular model with two change-points under the alternative hypothesis, is also studied. Numerical studies by Monte-Carlo simulations show the performance of the proposed test statistic, compared to an existing method in literature, for models without change or with one or two change-points.

**Keywords** Change-point · Nonlinear parametric model · Empirical likelihood test · Asymptotic behaviour.

## 1 Introduction

We consider a classical model of parametric nonlinear regression :

$$Y_i = f(\mathbf{X}_i; \boldsymbol{\beta}) + \varepsilon_i, \qquad i = 1, \cdots, n, \tag{1}$$

where a possible change in the regression parameters could occurs. This is called, change-point problem.

G.Ciuperca - Z. Salloum
Université de Lyon, Université Lyon 1, CNRS, UMR 5208, Institut Camille Jordan,
Bat. Braconnier, 43, blvd du 11 novembre 1918, F - 69622 Villeurbanne Cedex, France
E-mail: Gabriela.Ciuperca@univ-lyon1.fr

Z. Salloum
E-mail: salloum@math.univ-lyon1.fr

Change-point detection problems fall in two categories. The first type is *a posteriori*: after that the $n$ all observations are realized, we study if, a certain moment $k \in \{2, \cdots, n-1\}$, the model (parameter $\boldsymbol{\beta}$, to be more precise) is changed :

$$Y_i = \begin{cases} f(\mathbf{X}_i; \boldsymbol{\beta}_1) + \varepsilon_i & i = 1, \cdots, k \\ f(\mathbf{X}_i; \boldsymbol{\beta}_2) + \varepsilon_i & i = k+1, \cdots, n. \end{cases} \tag{2}$$

The second type of change-points model is sequential (*a priori*), where the change detection is performed in real time. If in the first $k-1$ observations no change in the parameter regression has occurred, at observation $k$ we test that there is no change in the model: $Y_i = f(X_i; \boldsymbol{\beta}) + \varepsilon_i$, for all $i = 1, \cdots, k$, against the hypothesis that the model has the form :

$$\begin{array}{ll} Y_i = f(\mathbf{X}_i; \boldsymbol{\beta}) + \varepsilon_i & \text{for } i = 1, \cdots, k-1 \\ Y_k = f(\mathbf{X}_k; \boldsymbol{\beta}^*) + \varepsilon_k, & \end{array} \tag{3}$$

with $\boldsymbol{\beta} \neq \boldsymbol{\beta}^*$.

In this paper, we consider a posteriori change-point problem.

For the two types of problems, the number of publications in the last years is every extensive. Let us mention some references concerning the sequential change-point problem. If the function $f$ is linear, $f(\mathbf{x}, \boldsymbol{\beta}) = \mathbf{x}^t \boldsymbol{\beta}$, in the papers [10], [11], the CUSUM method is used to find a test statistic for detecting the presence or absence of a change. The results have been generalized by [6] for a nonlinear model. We can also mention the papers [12], [15], [16] for the sequential detection of a change-point.

For a posteriori change-point problem, in order to detect a change-point presence, model (1) is tested against model (2). The non-identifiability of model under the null hypothesis makes classical test techniques unusable. In many articles in the literature, the authors propose criteria: see for example [17], [4], [21]. Various hypothesis tests have been proposed only for the linear models. The likelihood-ratio test method is used in [1] and [13]. A non-parametric approach based on Empirical Likelihood (EL) for testing a change in a linear model is considered by [14]. Always using the EL method, the papers [23], [22] construct the confidence region for the coefficient difference of a two-sample linear regression model. For a linear quantile model, [18] proposes two types of statistics: one based on the subgradient and an another based on Wald statistic.

In this paper, we consider the change-point problem in a general nonlinear model, by the EL method. Then, the framework of [14] is generalized. One of the major difficulties for nonlinear model (beside the linear model approach) is that, for finding the test statistic, the corresponding score functions depend on the regression parameters, and above all, the analytical form of these derivatives is unknown. On the other hand, for linear models, many proofs are based on the convexity of the regression function with respect to the parameter regression, then, the extreme value of a convex function is attained on the boundary. These two factors lead to a more difficult theoretical study of the test statistics for nonlinear model. Another difficulty to study the properties of the test statistic, for detecting a change in model, is due to the

dependence on the change-points of the regression parameter estimator. To the authors' knowledge, the only paper which studies a hypothesis test in a change-point nonlinear model is [3]for very smooth nonlinear functions, using the least square method. But the least square method, in respect to the EL method, has the disadvantage that is less efficient for outliers data. This occurs in the case of fatter tailed distributions of the error term. Moreover, we will see in Section 2 that the considered assumptions in [3] are stronger than in the present paper.
Recall also the paper [9] which tests the structural stability in a nonlinear model by a generalized method of moments, but where the alternative hypothesis is not a change in the regression parameters.
I would emphasize that in the present paper, we have obtained an interesting result concerning the numerical simulations. The EL test outperforms the change detection by least square(LS) test proposed by [3]. The LS test does not work when the change-point is off-centred in the measurement interval. The proposed EL test does not this defect.

The paper is organized as follows. We first construct in Section 2 a statistic, in order to test the change in the regression parameters of the nonlinear model. The asymptotic behaviour of the test statistic under the null hypothesis as well as under the alternative hypothesis is studied. A particular case of two change-point model, the epidemic model, is considered in Section 3. In Section 4, simulations results illustrate the performance of the proposed test, concerning the empirical size, the asymptotic power and the estimation of the time of change, in particular when the error distribution is not Gaussian, when it has outliers or a large standard deviation. Some lemmas and their proofs are given in the last section (Appendix, Section 5).

## 2 Test with one change-point

In this section, for a nonlinear model we are going to test the hypothesis that there is no change in the parameters of model (1) against the hypothesis that the parameters change from $\boldsymbol{\beta}_1$ to $\boldsymbol{\beta}_2$ at an unknown observation k (model (2)).

### 2.1 Model, notations, assumptions

For each observation $i$, $Y_i$ denotes the response variable, $\mathbf{X}_i$ is a $p \times 1$ random vector of regressors with distribution function $H(\mathbf{x})$, with $\mathbf{x} \in \Upsilon$, $\Upsilon \subseteq I\!\!R^p$, and $\varepsilon_i$ is the error.
The continuous random vector sequence $(\mathbf{X}_i, \varepsilon_i)_{1 \leq i \leq n}$ is independent identically distributed (i.i.d), with the same joint distribution as $(\mathbf{X}, \varepsilon)$. For all $i$, $\varepsilon_i$ is independent of $\mathbf{X}_i$.
The regression function $f : \Upsilon \times \Gamma \to \mathbb{R}$, with $\Upsilon \subseteq \mathbb{R}^p$, and $\Gamma \subseteq \mathbb{R}^d$, is known up

to a $d$-dimensional parameter $\boldsymbol{\beta}$. The parameter set $\Gamma$ is supposed compact. With regard to the random variable $\varepsilon$ we make following assumption :
**(A1)** $\mathbb{E}[\varepsilon_i] = 0$ and $\mathbb{E}[\varepsilon_i^2] < \infty$, for all $i = 1, \cdots, n$.
The regression function $f : \Upsilon \times \Gamma \to \mathbb{R}$ and the random vector $\mathbf{X}$ satisfy the conditions :
**(A2)** for all $\mathbf{x} \in \Upsilon$ and for $\boldsymbol{\beta} \in \Gamma$, the function $f(\mathbf{x}, \boldsymbol{\beta})$ is thrice differentiable in $\boldsymbol{\beta}$ and continuous on $\Upsilon$.
In following, for $\mathbf{x} \in \Upsilon$ and $\boldsymbol{\beta} \in \Gamma$, we use notation $\dot{\mathbf{f}}(\mathbf{x}, \boldsymbol{\beta}) \equiv \partial f(\mathbf{x}, \boldsymbol{\beta})/\partial \boldsymbol{\beta}$, $\ddot{\mathbf{f}}(\mathbf{x}, \boldsymbol{\beta}) \equiv \partial^2 f(\mathbf{x}, \boldsymbol{\beta})/\partial \boldsymbol{\beta}^2$ and $f^{(3)}(\mathbf{x}, \boldsymbol{\beta}) \equiv \partial^3 f(\mathbf{x}, \boldsymbol{\beta})/\partial \boldsymbol{\beta}^3$.
**(A3)** $\|\ddot{\mathbf{f}}(\mathbf{x}, \boldsymbol{\beta})\|_2$, $\|\mathbf{f}^{(3)}(\mathbf{x}, \boldsymbol{\beta})\|_2$ are bounded for any $\mathbf{x} \in \Upsilon$ and $\boldsymbol{\beta}$ in a neighbourhood of $\boldsymbol{\beta}^0$.
**(A4)** $\mathbb{E}[\dot{\mathbf{f}}(X, \boldsymbol{\beta})] < \infty$ and $\mathbb{E}[\dot{\mathbf{f}}(X, \boldsymbol{\beta})\dot{\mathbf{f}}^t(X, \boldsymbol{\beta})] < \infty$, for $\boldsymbol{\beta}$ in a neighbourhood of $\boldsymbol{\beta}^0$.

Assumptions (A3), (A4) are standard conditions, which are used in nonlinear models, for example see paper [19]. We remark that assumption (A4) is weaker than the corresponding assumption employed in paper [3], where the least square method is used to test $H_0$ against $H_1$ : $\sup_{\mathbf{x}, \boldsymbol{\beta}} \mathbb{E}[\dot{\mathbf{f}}(\mathbf{x}, \boldsymbol{\beta})]^{2s} < \infty$, $\sup_{\mathbf{x}, \boldsymbol{\beta}} \mathbb{E}[\ddot{\mathbf{f}}(\mathbf{x}, \boldsymbol{\beta})]^{2s} < \infty$, for some $s > 2$.

We are interested in testing of the null hypothesis of no change in the model (2). Then the model has the form (1), that is
$H_0$: $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \boldsymbol{\beta}$.
The alternative hypothesis assumes that one change occurs in the regression parameters, that is
$H_1$: $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$.
Let $\boldsymbol{\beta}^{\mathbf{0}}$ denote the true (unknown) of the parameter $\boldsymbol{\beta}$ under hypothesis $H_0$ and $\boldsymbol{\beta}_1^0$, $\boldsymbol{\beta}_2^0$ (also unknown) the true parameters under hypothesis $H_1$.

In addition to the notations introduced above, we define notations which will be used under hypothesis $H_0$. Let us consider the following $d$-random vectors $\mathbf{g}(\mathbf{X}_i, \boldsymbol{\beta}) \equiv \mathbf{g}_i(\boldsymbol{\beta}) \equiv \dot{\mathbf{f}}(\mathbf{X}_i, \boldsymbol{\beta})[Y_i - f(\mathbf{X}_i, \boldsymbol{\beta})]$. We remark that, under the hypothesis $H_0$, we have $\mathbf{g}_i(\boldsymbol{\beta}^{\mathbf{0}}) = \dot{\mathbf{f}}(\mathbf{X}_i, \boldsymbol{\beta}^{\mathbf{0}})\varepsilon_i$, for all $i = 1, \cdots, n$ and $\mathbb{E}[\mathbf{g}_i(\boldsymbol{\beta}^{\mathbf{0}})] = 0$. Consider also the $d \times d$ matrix $\mathbf{V} \equiv \mathbb{E}[\dot{\mathbf{f}}(\mathbf{X}_i, \boldsymbol{\beta}^{\mathbf{0}})\dot{\mathbf{f}}^t(\mathbf{X}_i, \boldsymbol{\beta}^{\mathbf{0}})]$. Then $\sigma^2 \mathbf{V} = \mathbb{V}\mathrm{ar}\,(\varepsilon_i \dot{\mathbf{f}}(\mathbf{X}_i, \boldsymbol{\beta}^{\mathbf{0}}))$.
Let $y_1, \cdots, y_k, y_{k+1}, \cdots, y_n$ be observations for the random variables $Y_1, \cdots, Y_k$, $Y_{k+1}, \cdots, Y_n$. Consider the following sets $I \equiv \{1, ..., k\}$ and $J \equiv \{k+1, ..., n\}$, which contain the observation subscripts of the two segments for the model (2). Corresponding to these sets, let be the probability vectors : $(p_1, \cdots, p_k)$ and $(q_{k+1}, \cdots, q_n)$. These vectors contained the probability to observe the value $y_i$ (respectively $y_j$) for the dependent variable $Y_i$ (respectively $Y_j$) : $p_i \equiv \mathbb{P}[Y_i = y_i]$, for $i = 1, \cdots, k$ and $q_j \equiv \mathbb{P}[Y_j = y_j]$, for $j = k+1, \cdots, n$. Obviously, $\sum_{i \in I} p_i = 1$, $\sum_{j \in J} q_j = 1$.

All throughout the paper, C denotes a positive generic constant which may take different values in different formula or even in different parts of the same formula. All vectors are column and $\mathbf{v}^t$ denotes the transposed of $\mathbf{v}$. All vectors and matrices are in bold. Concerning the used norms, for a m-vector $\mathbf{v} = (v_1, \cdots, v_m)$, let us denote by $\|\mathbf{v}\|_1 = \sum_{j=1}^m |v_j|$ its $L_1$-norm and $\|\mathbf{v}\|_2 = (\sum_{j=1}^m v_j^2)^{1/2}$ its $L_2$-norm. For a matrix $\mathbf{M} = (a_{ij})_{\substack{1\leqslant i\leqslant m_1\\ 1\leqslant j\leqslant m_2}}$, we denote by $\|\mathbf{M}\|_1 = \max_{j=1,\cdots,m_2}(\sum_{i=1}^{m_1} |a_{ij}|)$, the subordinate norm to the vector norm $\|.\|_1$. Let $\underset{n\to\infty}{\overset{\mathcal{L}}{\longrightarrow}}$, $\underset{n\to\infty}{\overset{I\!P}{\longrightarrow}}$, $\underset{n\to\infty}{\overset{a.s.}{\longrightarrow}}$ represent convergence in distribution, in probability and almost sure, respectively, as $n \to \infty$.
For coherence, we try to use the some notations as in the paper [14], where the linear model was considered. This will allow to highlight the difficulties and results due to the nonlinearity.

## 2.2 Test statistic

Under hypothesis $H_0$, the profile empirical likelihood (EL) for $\boldsymbol{\beta}$ is

$$\mathcal{R}_0(\boldsymbol{\beta}) = \sup_{(p_1,\cdots,p_k)} \sup_{(q_{k+1},\cdots,q_n)} \Big\{ \prod_{i\in I} p_i \prod_{j\in J} q_j; \sum_{i\in I} p_i = 1, \sum_{j\in J} q_j = 1, \\ \sum_{i\in I} p_i \mathbf{g}_i(\boldsymbol{\beta}) = \sum_{j\in J} q_j \mathbf{g}_j(\boldsymbol{\beta}) = \mathbf{0}_d \Big\},$$

with $\mathbf{0}_d$ the $d$-vector with all components zero. Without constraints $\sum_{i\in I} p_i \mathbf{g}_i(\boldsymbol{\beta}) = \mathbf{0}_d$, the maximum of $\prod_{i\in I} p_i$, $\prod_{j\in J} q_j$ are attained for $p_i = k^{-1}$, $q_j = (n-k)^{-1}$, respectively. Then, the profile EL ratio for $\boldsymbol{\beta}$ has the form

$$\mathcal{R}_0'(\boldsymbol{\beta}) = \sup_{(p_1,\cdots,p_k)} \sup_{(q_{k+1},\cdots,q_n)} \Big\{ \prod_{i\in I} kp_i \prod_{j\in J} (n-k) q_j; \sum_{i\in I} p_i = 1, \\ \sum_{j\in J} q_j = 1, \sum_{i\in I} p_i \mathbf{g}_i(\boldsymbol{\beta}) = \sum_{j\in J} q_j \mathbf{g}_j(\boldsymbol{\beta}) = \mathbf{0}_d \Big\}. \quad (4)$$

Similarly, under hypothesis $H_1$, the profile EL is

$$\mathcal{R}_1(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \sup_{(p_1,\cdots,p_k)} \sup_{(q_{k+1},\cdots,q_n)} \Big\{ \prod_{i\in I} p_i \prod_{j\in J} q_j; \sum_{i\in I} p_i = 1, \\ \sum_{j\in J} q_j = 1, \sum_{i\in I} p_i \mathbf{g}_i(\boldsymbol{\beta}_1) = \mathbf{0}_d, \sum_{j\in J} q_j \mathbf{g}_j(\boldsymbol{\beta}_2) = \mathbf{0}_d \Big\}.$$

Then, the profile EL ratio for $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ has the form

$$\mathcal{R}_1'(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \sup_{(p_1,\cdots,p_k)} \sup_{(q_{k+1},\cdots,q_n)} \Big\{ \prod_{i\in I} kp_i \prod_{j\in J} (n-k) q_j; \sum_{i\in I} p_i = 1, \\ \sum_{j\in J} q_j = 1, \sum_{i\in I} p_i \mathbf{g}_i(\boldsymbol{\beta}_1) = \mathbf{0}_d, \sum_{j\in J} q_j \mathbf{g}_j(\boldsymbol{\beta}_2) = \mathbf{0}_d \Big\}.$$

Thus, using an idea similar to the maximum likelihood test for testing $H_0$ against $H_1$, we consider the profile EL ratio

$$\frac{\mathcal{R}_0(\boldsymbol{\beta})}{\mathcal{R}_1(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)}=\frac{\mathcal{R}'_0(\boldsymbol{\beta})}{\mathcal{R}'_1(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)}, \tag{5}$$

but, under this form, it has a complicated expression. In order to find a simpler form for the test statistic, we will study the denominator behaviour of the process given by (5).

The following result is a generalization of the nonparametric version of the Wilks theorem. More specifically, under $H_1$ due to the observation independence, on each segment we have a Wilks theorem. The profile EL ratio for $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ has a $\chi^2$ asymptotic distribution.

**Theorem 1** *Suppose that assumptions (A1)-(A3) hold. Under the hypothesis $H_1$, we have*

$$-2\log\mathcal{R}'_1(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)\xrightarrow[n\to\infty]{\mathcal{L}}\chi^2(2d).$$

*Proof.* Under $H_1$, on the first segment generated by the observations for $i\in I$, the profile EL function for $\boldsymbol{\beta}_1$, for fixed $k$, is

$$\sup_{(p_1,\cdots,p_k)}\left\{\prod_{i\in I}kp_i; 0\leq p_i\leq 1, \sum_{i\in I}p_i=1, \sum_{i\in I}p_i\mathbf{g}_i(\boldsymbol{\beta}_1)=\mathbf{0}_d\right\}.$$

Using the Lagrange multiplier method, we consider the following random process $\sum_{i\in I}\log p_i+\eta(\sum_{i\in I}p_i-1)-k\boldsymbol{\lambda}_1^t(\sum_{i\in I}p_i\mathbf{g}_i(\boldsymbol{\beta}_1))$, with $\boldsymbol{\lambda}_1\in\mathbb{R}^d$, $\eta\in\mathbb{R}$. Taking derivative with respect to $p_i$ of this process equal to zero, we obtain

$$p_i=\frac{1}{k\boldsymbol{\lambda}_1^t\mathbf{g}_i(\boldsymbol{\beta}_1)-\eta}. \tag{6}$$

Then, $1+\eta p_i-kp_i\boldsymbol{\lambda}_1^t\mathbf{g}_i(\boldsymbol{\beta}_1)=0$, and summing for $i=1,\cdots,k$, we obtain that $\eta=-k$. Hence, the probability $p_i$ given by (6) becomes

$$p_i=\frac{1}{k(1+\boldsymbol{\lambda}_1^t\mathbf{g}_i(\boldsymbol{\beta}_1))}.$$

Similarly, the profile EL function on the second segment generated by the observations for $j\in J$, is

$$\sup_{(q_{k+1},\cdots,q_n)}\left\{\prod_{j\in J}(n-k)q_j; 0\leq q_j\leq 1, \sum_{j\in J}q_j=1, \sum_{j\in J}q_j\mathbf{g}_j(\boldsymbol{\beta}_2)=\mathbf{0}_d\right\}.$$

This function is maximed for $q_j=(n-k-\boldsymbol{\lambda}_2^t\mathbf{g}_i(\boldsymbol{\beta}_2))^{-1}$, with $\boldsymbol{\lambda}_2\in\mathbb{R}^p$. Then the empirical log-likelihood ratio statistic can be written

$$-2\log\mathcal{R}'_1(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)=2\sum_{i\in I}\log\left[1+\boldsymbol{\lambda}_1^t\mathbf{g}_i(\boldsymbol{\beta}_1)\right]+2\sum_{j\in J}\log\left[1-\boldsymbol{\lambda}_2^t\mathbf{g}_i(\boldsymbol{\beta}_2)\right]. \tag{7}$$

In view of Theorem 4.1 of [7], using assumptions (A1), (A2) and (A3), each sum of the right-hand side of (7) converges in law to $\chi^2(d)$. Taking into account that the two terms of relation (7) involved two independent sets of random vectors we obtain the theorem. □

Consequently of this theorem, the denominator of the EL ratio given by (5), is not asymptotically depend on the parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, under hypothesis $H_1$. Then, from now on, we are going to consider that test statistic $-2\log \mathcal{R}'_0(\boldsymbol{\beta})$.

Taking into account the expression of $\mathcal{R}'_0(\boldsymbol{\beta})$ given by (4), and using the Lagrange multiplier method, we have that maximizing $-2\log \mathcal{R}'_0(\boldsymbol{\beta})$ is equivalent to maximizing the following statistic with respect to $\boldsymbol{\beta}$, $\eta_1$, $\eta_2$, $\boldsymbol{\lambda}_1$, $\boldsymbol{\lambda}_2$,

$$\sum_{i\in I}\big[\log p_i - n\boldsymbol{\lambda}_1^t p_i \mathbf{g}_i(\boldsymbol{\beta})\big] + \sum_{j\in J}\big[\log q_j + n\boldsymbol{\lambda}_2^t q_j \mathbf{g}_j(\boldsymbol{\beta})\big] + \eta_1\big(\sum_{i\in I} p_i - 1\big) + \eta_2\big(\sum_{j\in J} q_j - 1\big) \tag{8}$$

where $\boldsymbol{\beta}\in \Gamma$, $\eta_1, \eta_2 \in \mathbb{R}$ and $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 \in \mathbb{R}^d$.
Since the derivatives of (8) with respect to $p_i$, $q_j$ are null, using a similar argument as in the proof of Theorem 1, we obtain that

$$p_i = \frac{1}{k + n\boldsymbol{\lambda}_1^t \mathbf{g}_i(\boldsymbol{\beta})} \qquad q_j = \frac{1}{n-k-n\boldsymbol{\lambda}_2^t \mathbf{g}_j(\boldsymbol{\beta})}. \tag{9}$$

Then, the statistic $-2\log \mathcal{R}'_0(\boldsymbol{\beta})$ becomes

$$2\sum_{i\in I}\log\big[1+\frac{n}{k}\boldsymbol{\lambda}_1^t \mathbf{g}_i(\boldsymbol{\beta})\big] + 2\sum_{j\in J}\log\big[1-\frac{n}{n-k}\boldsymbol{\lambda}_2^t \mathbf{g}_j(\boldsymbol{\beta})\big]. \tag{10}$$

Taking into account relation (9), for the probabilities $p_i$ and $q_j$, the derivative with respect to $\boldsymbol{\beta}$ of (10) is $2n\big[\sum_{i\in I} p_i \boldsymbol{\lambda}_i^t \dot{\mathbf{g}}_i(\boldsymbol{\beta}) - \sum_{j\in J} q_j \boldsymbol{\lambda}_2^t \dot{\mathbf{g}}_j(\boldsymbol{\beta})\big] = 0$, with $\dot{\mathbf{g}}_i(\boldsymbol{\beta})$ the $d\times d$ matrix of the derivatives of vector $\mathbf{g}_i(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$, for $i = 1, \cdots, k$. In order to have single parameters $\boldsymbol{\lambda}$, we restrict the study to a particular case, when $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ satisfy the constraint $\mathbf{V}_1(\boldsymbol{\beta})\boldsymbol{\lambda}_1 = \mathbf{V}_2(\boldsymbol{\beta})\boldsymbol{\lambda}_2$, with $\mathbf{V}_1(\boldsymbol{\beta}) \equiv k^{-1}\sum_{i\in I}\dot{\mathbf{g}}_i(\boldsymbol{\beta})$, $\mathbf{V}_2(\boldsymbol{\beta}) \equiv (n-k)^{-1}\sum_{j\in J}\dot{\mathbf{g}}_j(\boldsymbol{\beta})$. In the case of the true parameter $\boldsymbol{\beta}^{\mathbf{0}}$, this two last matrices become $\mathbf{V}_1^0 \equiv \mathbf{V}_1(\boldsymbol{\beta}^{\mathbf{0}})$ and $\mathbf{V}_2^0 \equiv \mathbf{V}_2(\boldsymbol{\beta}^{\mathbf{0}})$. Considering this constraint, statistic (10) becomes

$$2\sum_{i\in I}\log\big[1+\frac{n}{k}\boldsymbol{\lambda}^t \mathbf{g}_i(\boldsymbol{\beta})\big] + 2\sum_{j\in J}\log\big[1-\frac{n}{n-k}\boldsymbol{\lambda}^t \mathbf{V}_1(\boldsymbol{\beta})\mathbf{V}_2^{-1}(\boldsymbol{\beta})\mathbf{g}_j(\boldsymbol{\beta})\big]. \tag{11}$$

In order that the parameters belong a bounded set, in the place of $k$, we consider $\theta_{nk} \equiv k/n$, and we denote statistic (11) by $Z_{nk}(\theta_{nk}, \boldsymbol{\lambda}, \boldsymbol{\beta})$. Under hypothesis $H_1$, if $k_0$ is the point where the model change, we denote $\theta_{0n} = k_0/n$. Similar to the classical maximum likelihood test, but for models without change-points, we will study the maximum of empirical log-likelihood test statistic. Then, we calculate the score functions of test statistic (11)

$$\begin{aligned}\boldsymbol{\phi}_{1n}(\theta_{nk},\boldsymbol{\lambda},\boldsymbol{\beta}) &\equiv \frac{\partial Z_{nk}(\theta_{nk},\boldsymbol{\lambda},\boldsymbol{\beta})}{2\partial\boldsymbol{\lambda}} \\ &= \sum_{i\in I}\frac{\mathbf{g}_i(\boldsymbol{\beta})}{\theta_{nk}+\boldsymbol{\lambda}^t\mathbf{g}_i(\boldsymbol{\beta})} - \sum_{j\in J}\frac{\mathbf{V}_1(\boldsymbol{\beta})\mathbf{V}_2^{-1}(\boldsymbol{\beta})\mathbf{g}_j(\boldsymbol{\beta})}{1-\theta_{nk}-\boldsymbol{\lambda}^t\mathbf{V}_1(\boldsymbol{\beta})\mathbf{V}_2^{-1}(\boldsymbol{\beta})\mathbf{g}_j(\boldsymbol{\beta})}.\end{aligned} \tag{12}$$

$$\begin{aligned}\boldsymbol{\phi}_{2n}(\theta_{nk},\boldsymbol{\lambda},\beta) &\equiv \frac{\partial Z_{nk}(\theta_{nk},\boldsymbol{\lambda},\boldsymbol{\beta})}{2\partial\boldsymbol{\beta}} \\ &= \sum_{i\in I}\frac{\dot{\mathbf{g}}_i(\boldsymbol{\beta})\boldsymbol{\lambda}}{\theta_{nk}+\boldsymbol{\lambda}^t(\boldsymbol{\beta})\mathbf{g}_i(\boldsymbol{\beta})} \\ &- \sum_{j\in J}\frac{\boldsymbol{\lambda}^t(\boldsymbol{\beta})\dot{\mathbf{V}}_1(\boldsymbol{\beta})\mathbf{V}_2(\boldsymbol{\beta})^{-1}\mathbf{g}_j(\boldsymbol{\beta})+\boldsymbol{\lambda}^t(\boldsymbol{\beta})\mathbf{V}_1(\boldsymbol{\beta})\dot{\mathbf{V}}_2^{-1}(\boldsymbol{\beta})\mathbf{g}_j(\boldsymbol{\beta})}{1-\theta_{nk}-\boldsymbol{\lambda}^t\mathbf{V}_1(\boldsymbol{\beta})\mathbf{V}_2^{-1}(\boldsymbol{\beta})\mathbf{g}_j(\boldsymbol{\beta})} \\ &- \sum_{j\in J}\frac{(\boldsymbol{\beta})\mathbf{V}_1(\boldsymbol{\beta})\mathbf{V}_2^{-1}(\boldsymbol{\beta})\dot{\mathbf{g}}_j(\boldsymbol{\beta})\boldsymbol{\lambda}}{1-\theta_{nk}-\boldsymbol{\lambda}^t\mathbf{V}_1(\boldsymbol{\beta})\mathbf{V}_2^{-1}(\boldsymbol{\beta})\mathbf{g}_j(\boldsymbol{\beta})}.\end{aligned} \tag{13}$$

Then, solving the system $\boldsymbol{\phi}_{1n}(\theta_{nk},\boldsymbol{\lambda},\boldsymbol{\beta})=\mathbf{0}_d$ and $\boldsymbol{\phi}_{2n}(\theta_{nk},\boldsymbol{\lambda},\boldsymbol{\beta})=\mathbf{0}_d$, the obtained solutions $\hat{\boldsymbol{\lambda}}(\theta_{nk})$ and $\hat{\boldsymbol{\beta}}(\theta_{nk})$ are the maximizers of the statistic (11). We so obtain the profile maximum empirical likelihood function $Z_{nk}(\theta_{nk},\hat{\boldsymbol{\lambda}}(\theta_{nk}),\hat{\boldsymbol{\beta}}(\theta_{nk}))$, which depends only on the change-point parameter $\theta_{nk}$.
We emphasise that, compared with a linear model, in our case, matrix $\mathbf{V}_1(\boldsymbol{\beta})$, $\mathbf{V}_2(\boldsymbol{\beta})$ and derivative $\dot{\mathbf{g}}(\boldsymbol{\beta})$ depend on $\boldsymbol{\beta}$. These, besides the nonlinearity of $\mathbf{g}(\boldsymbol{\beta})$ involve difficulties in the study of the statistic $Z_{nk}(\theta_{nk},\boldsymbol{\lambda},\boldsymbol{\beta})$ and of the solutions $\hat{\boldsymbol{\lambda}}(\theta_{nk})$, $\hat{\boldsymbol{\beta}}(\theta_{nk})$.

## 2.3 Asymptotic behaviour of the test statistic

In this section, for the probabilities given by (9), under the constraint $\mathbf{V}_1(\boldsymbol{\beta})\boldsymbol{\lambda}_1=\mathbf{V}_2(\boldsymbol{\beta})\boldsymbol{\lambda}_2$, we will first prove that $kp_i$, $(n-k)q_j$, can be framed by two strictly positive constants. This implies that the test statistic $Z_{nk}(\theta_{nk},\hat{\boldsymbol{\lambda}}(\theta_{nk}),\hat{\boldsymbol{\beta}}(\theta_{nk}))$ is well defined.
Properties established for $\hat{\boldsymbol{\lambda}}(\theta_{nk})$ and $\hat{\boldsymbol{\beta}}(\theta_{nk})$ will allow to consider instead of (11), a more simple statistical test, given by relation (25). Next, we will study the asymptotic behaviour of this statistic, firstly under the hypothesis $H_0$ and next under $H_1$.

### *2.3.1 Asymptotic behaviour under $H_0$*

We will first study $kp_i$, for $i\in I$, and $(n-k)q_j$, for $j\in J$.

**Proposition 1** *Let the $\eta$-neighbourhood of $\boldsymbol{\beta}^{\mathbf{0}}$, $\mathcal{V}_\eta(\boldsymbol{\beta}^{\mathbf{0}}) = \{\boldsymbol{\beta} \in \Gamma; \|\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}\|_2 \leq \eta\}$, with $\eta \to 0$. Under hypothesis $H_0$, suppose that assumptions (A1)-(A4) hold. Then we have*
*(i) For all $i \in I$, for all $\epsilon > 0$, there exist two constants $M_1, M_2 > 0$, such that, for all $\boldsymbol{\beta} \in \mathcal{V}_\eta(\boldsymbol{\beta}^{\mathbf{0}})$,*

$$I\!P\Big[\frac{1}{M_2} \leq \frac{1}{1+\frac{\boldsymbol{\lambda}^t}{\theta_{nk}} g_i(\boldsymbol{\beta})} \leq \frac{1}{M_1}\Big] \geq 1-\epsilon. \tag{14}$$

*(ii) For all $j \in J$, for all $\epsilon > 0$, there exist two constants $M_3, M_4 > 0$, such that, for all $\boldsymbol{\beta} \in \mathcal{V}_\eta(\boldsymbol{\beta}^{\mathbf{0}})$,*

$$I\!P\Big[\frac{1}{M_4} \leq \frac{1}{1-\frac{\boldsymbol{\lambda}^t}{1-\theta_{nk}} \mathbf{V}_1(\boldsymbol{\beta})(\mathbf{V}_2(\boldsymbol{\beta}))^{-1}\mathbf{g}_j(\boldsymbol{\beta})} \leq \frac{1}{M_3}\Big] \geq 1-\epsilon. \tag{15}$$

*Proof. (i)* We consider the following decomposition for the Lagrange multiplier: $\boldsymbol{\lambda} = \rho\boldsymbol{\phi}$, such that $\rho \geq 0$ and $\|\boldsymbol{\phi}\|_1 = 1$. Lemma 2 implies that, there exists $M_2 > 0$, such that

$$\frac{1}{1+\frac{\boldsymbol{\lambda}^t}{\theta_{nk}}\mathbf{g}_i(\boldsymbol{\beta})} \geq \frac{1}{1+\frac{\rho}{\theta_{nk}}\|\boldsymbol{\phi}^t\mathbf{g}_i(\boldsymbol{\beta})\|_1} \geq \frac{1}{1+\frac{\rho}{\theta_{nk}}\|\mathbf{g}_i(\boldsymbol{\beta})\|_1} \geq \frac{1}{M_2},$$

with probability close to 1, that is, for all $\epsilon > 0$,

$$I\!P[\frac{1}{1+\frac{\rho}{\theta_{nk}}\|\mathbf{g}_i(\boldsymbol{\beta})\|_1} \geq \frac{1}{M_2}] \geq 1-\frac{\epsilon}{2}. \tag{16}$$

For the right-hand side of relation (14), we assume the contrary, that is, there exists $M_1 > 0$ such that

$$\sup_{i\in I,\boldsymbol{\beta}\in\Gamma} \frac{1}{1+\frac{\boldsymbol{\lambda}^t}{\theta_{nk}}\mathbf{g}_i(\boldsymbol{\beta})} \geq \frac{1}{M_1}.$$

This is equivalent to the fact that there exists $\tilde{M} > 0$, such that

$$\inf_{i\in I,\boldsymbol{\beta}\in\Gamma} \frac{\boldsymbol{\lambda}^t}{\theta_{nk}}\mathbf{g}_i(\boldsymbol{\beta}) \leq -\tilde{M}.$$

Since $\boldsymbol{\lambda} = \rho\boldsymbol{\phi}$, $\rho > 0$, and $0 < \theta_{nk} < 1$, therefore exists $\tilde{\tilde{M}} > 0$ such that

$$\inf_{i\in I,\boldsymbol{\beta}\in\Gamma} \boldsymbol{\phi}^t\mathbf{g}_i(\boldsymbol{\beta}) \leq -\tilde{\tilde{M}}. \tag{17}$$

On the other hand, we have that $\inf_{i\in I,\boldsymbol{\beta}\in\Gamma} \boldsymbol{\phi}^t\mathbf{g}_i(\boldsymbol{\beta}) \geq -\inf_{i\in I,\boldsymbol{\beta}\in\Gamma} \|\mathbf{g}_i(\boldsymbol{\beta})\|_1$, with probability 1. Taking into account relation (17), there exists $\tilde{\tilde{M}} > 0$ such as $-\inf_{i\in I,\boldsymbol{\beta}\in\Gamma} \|\mathbf{g}_i(\boldsymbol{\beta})\|_1 \leq -\tilde{\tilde{M}}$ again too $\sup_{i\in I,\boldsymbol{\beta}\in\Gamma} \|\mathbf{g}_i(\boldsymbol{\beta})\|_1 \geq \tilde{\tilde{M}}$, which is in contradiction with relation ( 2). Then, the relation (14) holds.
*(ii)* Relation (15) can be proved in a similar way. □

By the following result, we show that $\hat{\boldsymbol{\lambda}}(\theta_{nk})$ and $\hat{\boldsymbol{\beta}}(\theta_{nk})$, the solutions of the score equations $\boldsymbol{\phi}_{1n}(\theta_{nk}, \boldsymbol{\lambda}, \boldsymbol{\beta}) = \mathbf{0}_d$ and $\boldsymbol{\phi}_{2n}(\theta_{nk}, \boldsymbol{\lambda}, \boldsymbol{\beta}) = \mathbf{0}_d$, have suitable properties. More precisely, we show that $\|\hat{\boldsymbol{\lambda}}(\theta_{nk})\|_2 \to 0$, as $n \to \infty$ and that $\hat{\boldsymbol{\beta}}(\theta_{nk})$ is a consistent estimator of $\boldsymbol{\beta}^0$, under hypothesis $H_0$. We also obtain their convergence rate.

**Theorem 2** *Suppose that the assumptions (A1)-(A4) hold. Under the hypothesis $H_0$, we have $\hat{\boldsymbol{\lambda}}(\theta_{nk}) = min\ \{\theta_{nk}, 1-\theta_{nk}\} O_{I\!P}((n\ min\ \{\theta_{nk}, 1-\theta_{nk}\})^{-1/2})$ and $\hat{\boldsymbol{\beta}}(\theta_{nk}) - \boldsymbol{\beta}^0 = O_{I\!P}((n\ min\ \{\theta_{nk}, 1-\theta_{nk}\})^{-1/2})$.*

*Proof.* The structure of the proof is similar to that of linear model (Lemma A1 of [14]) but important modifications and supplementary results are necessary, due to the model nonlinearity. Without loss of generality, we assume that $\min\{\theta_{nk}, 1-\theta_{nk}\} = \theta_{nk}$. The other case is similar.
By the definition of the profile empirical likelihood ratio $\mathcal{R}'_0(\boldsymbol{\beta})$, we have the following constraints

$$\mathbf{0}_d = \sum_{i \in I} p_i \mathbf{g}_i(\boldsymbol{\beta}) = \sum_{j \in J} q_j \mathbf{g}_j(\boldsymbol{\beta}). \tag{18}$$

We recall that, under hypothesis $H_0$, the expression of $p_i$ is given by (9), and it is equal to $(\theta_{nk} + n\boldsymbol{\lambda}^t \mathbf{g}_i(\boldsymbol{\beta}))^{-1}$, for $i = 1, \cdots, n\theta_{nk}$. Then, by elementary calculations, we obtain

$$\mathbf{0}_d = \frac{1}{n\theta_{nk}} \sum_{i \in I} \mathbf{g}_i(\boldsymbol{\beta}) - \frac{1}{n\theta_{nk}^2} \sum_{i \in I} \frac{\mathbf{g}_i(\boldsymbol{\beta})\mathbf{g}_i^t(\boldsymbol{\beta})}{1 + \frac{\boldsymbol{\lambda}^t(\boldsymbol{\beta})}{\theta_{nk}} \mathbf{g}_i(\boldsymbol{\beta})} \boldsymbol{\lambda}(\boldsymbol{\beta}). \tag{19}$$

Let us make the remark that we denote $\boldsymbol{\lambda}$ by $\boldsymbol{\lambda}(\boldsymbol{\beta})$ in order to indicate that for each value of $\boldsymbol{\beta}$, solution of (19), we will have a different value for $\boldsymbol{\lambda}$. We take $\boldsymbol{\beta} = \boldsymbol{\beta}^0 \pm (n\theta_{nk})^{-r} \mathbf{1}_d$, with $\mathbf{1}_d$ the $d$-vector with all components 1 and $r > 0$ will be specified later. Therefore, $\|\boldsymbol{\beta} - \boldsymbol{\beta}^{\mathbf{0}}\|_2 = (n\theta_{nk})^{-r} \to 0$, as $n\theta_{nk} \to \infty$.

For the first sum of the right-hand side of (19), by Lemma 3, we have

$$\frac{1}{n\theta_{nk}} \sum_{i \in I} \mathbf{g}_i(\boldsymbol{\beta}) = O_{I\!P}((n\theta_{nk})^{-1/2}) + \mathbf{V}_1^0(\boldsymbol{\beta} - \boldsymbol{\beta}^{\mathbf{0}}) + o_{I\!P}(\boldsymbol{\beta} - \boldsymbol{\beta}^0).$$

Now, we consider the second term of the right-hand side of relation (19). From Proposition 1, we have that for all $\epsilon > 0$, there exists $M_1, M_2 > 0$, such that

$$I\!P\Big[\frac{1}{M_1} \sum_{i \in I} \mathbf{g}_i(\boldsymbol{\beta})\mathbf{g}_i^t(\boldsymbol{\beta}) \leq \sum_{i \in I} \frac{\mathbf{g}_i(\boldsymbol{\beta})\mathbf{g}_i^t(\boldsymbol{\beta})}{1 + \frac{\boldsymbol{\lambda}^t_(\boldsymbol{\beta})}{\theta_{nk}} \mathbf{g}_i(\boldsymbol{\beta})} \leq \frac{1}{M_2} \sum_{i \in I} \mathbf{g}_i(\boldsymbol{\beta})\mathbf{g}_i^t(\boldsymbol{\beta})\Big] < \epsilon.$$

This implies that, in order to study the second term of the right-hand side of the relation (19), we must study only $(n\theta_{nk})^{-1} \sum_{i \in I} \mathbf{g}_i(\boldsymbol{\beta})\mathbf{g}_i^t(\boldsymbol{\beta})$. By a Taylor's expansion of $\mathbf{g}_i(\boldsymbol{\beta})$ in a neighbourhood of $\boldsymbol{\beta}^{\mathbf{0}}$, using an argument similar to the one used for the first term of (19), together with the assumption (A3), we obtain

$$\frac{1}{n\theta_{nk}} \sum_{i\in I} \mathbf{g}_i(\boldsymbol{\beta})\mathbf{g}_i^t(\boldsymbol{\beta}) = \frac{1}{n\theta_{nk}} \sum_{i\in I} \mathbf{g}_i(\boldsymbol{\beta^0})\mathbf{g}_i^t(\boldsymbol{\beta^0})(1+o_{I\!P}(1)). \tag{20}$$

Taking into account Lemma 3 and relation (20), the relation (19) becomes

$$\mathbf{0}_d = \Big[O_{I\!P}((n\theta_{nk})^{-1/2}) + \mathbf{V}_1^0(\boldsymbol{\beta}-\boldsymbol{\beta^0}) - \frac{1}{n\theta_{nk}^2} \sum_{i=1}^{n\theta_{nk}} \mathbf{g}_i(\boldsymbol{\beta^0})\mathbf{g}_i^t(\boldsymbol{\beta^0})\boldsymbol{\lambda}(\boldsymbol{\beta})\Big](1+o_{I\!P}(1)) \tag{21}$$

We consider a constant $r$ such that $1/3 \le r < 1/2$. If $\boldsymbol{\beta} = \boldsymbol{\beta^0} + (n\theta_{nk})^{-r}\mathbf{1}_d$, then $(\boldsymbol{\beta}-\boldsymbol{\beta^0})^t\mathbf{1}_d > 0$, and if $\boldsymbol{\beta} = \boldsymbol{\beta^0} - (n\theta_{nk})^{-r}\mathbf{1}_d$ then $(\boldsymbol{\beta}-\boldsymbol{\beta^0})^t\mathbf{1}_d < 0$. Then, the relation (21) implies

$$\begin{aligned}\boldsymbol{\lambda}(\boldsymbol{\beta^0} \pm (n\theta_{nk})^{-r}\mathbf{1}_d) = \pm\Big[\theta_{nk}\Big(\frac{1}{n\theta_{nk}} \sum_{i\in I} \varepsilon_i^2 \dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})\Big)^{-1} \mathbf{V}_1^0 (n\theta_{nk})^{-r}\mathbf{1}_d \\ + O_{I\!P}((n\theta_{nk})^{-1/2})\Big](1+o_{I\!P}(1)).\end{aligned} \tag{22}$$

For the observations $j \in J$, let us consider the function $\mathbf{v} : \Gamma \to I\!R^d$ defined by

$$\mathbf{v}(\boldsymbol{\beta}) = \sum_{j\in J} q_j \mathbf{g}_j(\boldsymbol{\beta}) = \frac{1}{n-n\theta_{nk}} \sum_{j\in J} \frac{\mathbf{g}_j(\boldsymbol{\beta})}{1 - \frac{\boldsymbol{\lambda}^t(\boldsymbol{\beta})}{1-\theta_{nk}} \mathbf{V}_1(\boldsymbol{\beta})\mathbf{V}_2^{-1}(\boldsymbol{\beta})\mathbf{g}_j(\boldsymbol{\beta})}.$$

Note that $\mathbf{v}(\hat{\boldsymbol{\beta}}(\theta_{nk})) = \mathbf{0}_d$. For $\mathbf{v}(\boldsymbol{\beta})$, we have the following decomposition

$$\frac{\mathbf{V}_1(\boldsymbol{\beta})\mathbf{V}_2^{-1}(\boldsymbol{\beta})}{n(1-\theta_{nk})^2} \sum_{j\in J} \frac{\mathbf{g}_j(\boldsymbol{\beta})\mathbf{g}_j^t(\boldsymbol{\beta})}{1 - \frac{\boldsymbol{\lambda}^t(\boldsymbol{\beta})}{1-\theta_{nk}} \mathbf{V}_1(\boldsymbol{\beta})\mathbf{V}_2^{-1}(\boldsymbol{\beta})\mathbf{g}_j(\boldsymbol{\beta})} \boldsymbol{\lambda}(\boldsymbol{\beta}) + \frac{1}{n(1-\theta_{nk})} \sum_{j\in J} \mathbf{g}_j(\boldsymbol{\beta}).$$

To facilitate writing, we consider the following $d \times d$ squares matrices, defined by

$$\mathbf{D}_1^0 = \frac{1}{n\theta_{nk}} \sum_{i\in I} \mathbf{g}_i(\boldsymbol{\beta^0})\mathbf{g}_i^t(\boldsymbol{\beta^0}), \quad \mathbf{D}_2^0 = \frac{1}{n-n\theta_{nk}} \sum_{j\in J} \mathbf{g}_j(\boldsymbol{\beta^0})\mathbf{g}_j^t(\boldsymbol{\beta^0}). \tag{23}$$

As for the observations $i \in I$, we obtain, similarly as for relation (21), $\mathbf{v}(\boldsymbol{\beta}) = \Big[\mathbf{V}_2^0(\boldsymbol{\beta}-\boldsymbol{\beta^0}) + \frac{1}{1-\theta_{nk}}\mathbf{V}_1^0(\mathbf{V}_2^0)^{-1}\mathbf{D}_2^0\boldsymbol{\lambda}(\boldsymbol{\beta}) + O_{I\!P}((n(1-\theta_{nk}))^{-1/2})\Big](1+o_{I\!P}(1))$. Replacing $\boldsymbol{\lambda}(\boldsymbol{\beta})$ by the value obtained in (22), we obtain $\mathbf{v}(\boldsymbol{\beta}) = [\mathbf{V}_2^0(\boldsymbol{\beta}-\boldsymbol{\beta^0}) + (\theta_{nk})(1-\theta_{nk})^{-1}\mathbf{V}_1^0(\mathbf{V}_2^0)^{-1}\mathbf{D}_2^0(\mathbf{D}_1^0)^{-1}\mathbf{V}_1^0(\boldsymbol{\beta}-\boldsymbol{\beta^0}) + O_{I\!P}((n(1-\theta_{nk}))^{-1/2}) + O_{I\!P}((n\theta_{nk})^{-1/2})](1+o_{I\!P}(1))$. Because $\boldsymbol{\beta} = \boldsymbol{\beta^0} \pm (n\theta_{nk})^{-r}\mathbf{1}_d$, $1/3 \le r < 1/2$ and $\min\{\theta_{nk}, 1-\theta_{nk}\} = \theta_{nk}$, then $v(\boldsymbol{\beta})$ becomes

$$[(\mathbf{V}_2^0 + \frac{\theta_{nk}}{1-\theta_{nk}}\mathbf{V}_1^0(\mathbf{V}_2^0)^{-1}\mathbf{D}_2^0(\mathbf{D}_1^0)^{-1}\mathbf{V}_1^0)(\boldsymbol{\beta}-\boldsymbol{\beta^0}) + O_{I\!P}((n\theta_{nk})^{-1/2}))](1+o_{I\!P}(1)). \tag{24}$$

This implies that $\mathbf{v}(\boldsymbol{\beta}^{\mathbf{0}} + (n\theta_{nk})^{-r}\mathbf{1}_d)$ and $\mathbf{v}(\boldsymbol{\beta}^{\mathbf{0}} - (n\theta_{nk})^{-r}\mathbf{1}_d)$ have a different signs, component by component. Moreover, because $\mathbf{v}$ contains continuous functions in the neighbourhood of $\boldsymbol{\beta}^{\mathbf{0}}$, there exists a $\boldsymbol{\beta}$ such that $\mathbf{v}(\boldsymbol{\beta}) = \mathbf{0}_d$. But $\mathbf{v}(\hat{\boldsymbol{\beta}}(\theta_{nk})) = \mathbf{0}_d$, then $\hat{\boldsymbol{\beta}}(\theta_{nk}) \in [\boldsymbol{\beta}^{\mathbf{0}} - (n\theta_{nk})^{-r}\mathbf{1}_d, \boldsymbol{\beta}^{\mathbf{0}} + (n\theta_{nk})^{-r}\mathbf{1}_d]$, which implies, because $r < 1/2$, that $\hat{\boldsymbol{\beta}}(\theta_{nk}) - \boldsymbol{\beta}^{\mathbf{0}} = O_{I\!P}((n\theta_{nk})^{-r}) \geq O_{I\!P}((n\theta_{nk})^{-1/2})$. This last relation, together with the relation (24), since $\hat{\boldsymbol{\beta}}(\theta_{nk}) - \boldsymbol{\beta}^{\mathbf{0}}$ is the coefficient of a matrix strictly positive, implies that in order to have $\mathbf{v}(\hat{\boldsymbol{\beta}}(\theta_{nk})) = \mathbf{0}_d$, we must have $\hat{\boldsymbol{\beta}}(\theta_{nk}) - \boldsymbol{\beta}^{\mathbf{0}} = O_{I\!P}((n\theta_{nk})^{-1/2})$. Considering this result, for the relation (22), we obtain $\boldsymbol{\lambda}(\hat{\boldsymbol{\beta}}(\theta_{nk})) = \theta_{nk} O_{I\!P}((n\theta_{nk})^{-1/2})$. The theorem is completely proved. □

**Remark 1** *In view of the proof of Theorem 2, under hypothesis $H_0$, we can consider instead of $Z_{nk}(\theta_{nk}, \boldsymbol{\lambda}, \boldsymbol{\beta})$, given by (11), the following test statistic*

$$T_{nk}(\theta_{nk}, \boldsymbol{\lambda}, \boldsymbol{\beta}) = 2\sum_{i\in I} \log(1 + \frac{1}{\theta_{nk}}\boldsymbol{\lambda}^t \mathbf{g}_i(\boldsymbol{\beta})) + 2\sum_{j\in J} \log(1 - \frac{1}{1-\theta_{nk}}\boldsymbol{\lambda}^t \mathbf{g}_j(\boldsymbol{\beta})). \tag{25}$$

Because the regression function is nonlinear, and in order to the maximum empirical likelihood always exists, we consider that the parameter $\theta_{nk} \in [\Theta_{1n}, \Theta_{2n}] \subset (0,1)$, such that $n\Theta_{1n} \to \infty$, $n(1-\Theta_{2n}) \to \infty$, as $n \to \infty$ for example. The reader can find a discussion concerning the possible values of $\Theta_{1n}$, $\Theta_{2n}$ in the papers [24], [14]. Finally, the test statistic for testing the hypothesis $H_0$ against $H_1$ is

$$\tilde{T}_n \equiv \max_{\theta_{nk}\in[\Theta_{1n},\Theta_{2n}]} T_{nk}(\theta_{nk}, \hat{\boldsymbol{\lambda}}(\theta_{nk}), \hat{\boldsymbol{\beta}}(\theta_{nk})). \tag{26}$$

Then, we can consider as estimator for the time of change $k^0$, the maximum empirical likelihood estimator: $\tilde{k}_n \equiv n\tilde{\theta}_n \equiv n \min\{\tilde{\theta}_{nk}; \tilde{\theta}_{nk} = \arg\max_{\theta_{nk}\in[\Theta_{1n},\Theta_{2n}]} T_{nk}(\theta_{nk}, \hat{\boldsymbol{\lambda}}(\theta_{nk}), \hat{\boldsymbol{\beta}}(\theta_{nk}))\}$. Recall that $\hat{\boldsymbol{\lambda}}(\theta_{nk})$ and $\hat{\boldsymbol{\beta}}(\theta_{nk})$ are the solutions of the score equations (12) and (13).

The following result gives the asymptotic distribution of the test statistic $\tilde{T}_n$ given by (26), under the null hypothesis of no-change. For this purpose, we consider functions: $A(x) \equiv (2\log x)^{1/2}$, $D(x) = 2\log x + \log\log x$ and $u(n) = \frac{1-\Theta_{1n}\Theta_{2n}}{\Theta_{1n}(1-\Theta_{2n})} \to \infty$ as $n \to \infty$.

**Theorem 3** *Under the assumptions (A1)-(A4), if the hypothesis $H_0$ is true, then we have, for all $t \in \mathbb{R}$*

$$\lim_{n\to\infty} I\!P\{A(\log u(n))(\tilde{T}_n^{\frac{1}{2}} \leq t + D(\log u(n))\} = exp(-e^{-t}). \tag{27}$$

*Proof.* The proof is similar to that of Theorem 1.3.1 of [8], combining Theorem A.3.4 of [8] with Lemma 5. □

**Corollary 1** *Consequence of this theorem, for a fixed size* $\alpha \in (0,1)$*, we can deduct the critical test region :*

$$(\tilde{T}_n)^{1/2} \geq \frac{-\log(-\log \alpha) + D(\log u(n))}{A(\log u(n))}.$$

Using this result to applications is quite complicated; we must first solve equation system (12) and (13) where the nonlinearity in parameter $\boldsymbol{\beta}$ up to and including in matrices $\mathbf{V}_1(\boldsymbol{\beta})$, $\mathbf{V}_2(\boldsymbol{\beta})$, $\mathbf{V}_2^{-1}(\boldsymbol{\beta})$ causes numerical difficulties and long computation time. Moreover, it must then find $\theta_{nk}$ that maximizes statistic (26). We can propose an approached form for the test statistic much simpler to use in practice, but which preserves the theoretical properties of (26).

**Remark 2** *Taking into account the last relation of Lemma 5, Theorem 3 implies that, in practice, for testing the hypothesis* $H_0$ *against* $H_1$*, we will use an approximate form*

$$T(\theta_{nk}) = \Big(n\sigma^{-2}\theta_{nk}(1-\theta_{nk})(\overline{\boldsymbol{w}}_1 - \overline{\boldsymbol{w}}_2)^t \mathbf{V}^{-1}(\overline{\boldsymbol{w}}_1 - \overline{\boldsymbol{w}}_2)\Big)\Big(1 + o_{I\!P}(1)\Big), \quad (28)$$

*where :* $\overline{\boldsymbol{w}}_1 = (n\theta_{nk})^{-1}\sum_{i\in I}\mathbf{g}_i(\boldsymbol{\beta}^{\mathbf{0}})$ *and* $\overline{\boldsymbol{w}}_2 = (n(1-\theta_{nk}))^{-1}\mathbf{V}_1^0(\mathbf{V}_2^0)^{-1}\sum_{j\in J}\mathbf{g}_j(\boldsymbol{\beta}^{\mathbf{0}})$.

*Under* $H_0$*, error variance* $\sigma^2$ *is estimated by* $n^{-1}\sum_{i=1}^n[\boldsymbol{Y}_i - f(\mathbf{X}_i, \hat{\boldsymbol{\beta}}(\theta_{nk}))]^2$ *and matrix* $\mathbf{V}$ *by* $n^{-1}\sum_{i=1}^n \dot{\mathbf{f}}(\mathbf{X}_i, \hat{\boldsymbol{\beta}}(\theta_{nk}))\dot{\mathbf{f}}^t(\mathbf{X}_i, \hat{\boldsymbol{\beta}}(\theta_{nk}))$*. The approached maximum empirical likelihood estimator for the time of change* $k^0$ *is* $\hat{k}_n = n\hat{\theta}_n = n\min\{\hat{\theta}_{nk}; \hat{\theta}_{nk} = \arg\max_{\Theta_{1n}\leq\theta_{nk}\leq\Theta_{2n}} T(\theta_{nk})\}$.

### *2.3.2 Asymptotic behaviour under* $H_1$

We consider now that the hypothesis $H_1$ is true. If $k^0$ is the true time of change, we denote by $\theta_{n0} = k^0/n$ and we suppose that $\theta_0 \equiv \lim_{n\to\infty}\theta_{n0}$.
For $\mathbf{x} \in \Upsilon$ and $e \in I\!R$, let $F(\mathbf{x}, e)$ and $G(\mathbf{x}, e)$ the conditional distributions of $\mathbf{g}(\mathbf{X}_i, \boldsymbol{\beta})$ when $\mathbf{X}_i = \mathbf{x}$ for $i \in I$ and $j \in J$, respectively. Let $1\!\!1_{(.)}$ the indicator function. Recall that, the distribution function of $\mathbf{X}$ is $H(\mathbf{x})$. For $\mathbf{x}$ and $\theta$ fixed, we define
$dP(\mathbf{x}, e) \equiv (\theta 1\!\!1_{\{\theta\leq\theta_0\}} + \theta_0 1\!\!1_{\{\theta>\theta_0\}})dF(\mathbf{x}, e) + (\theta-\theta_0)1\!\!1_{\{\theta>\theta_0\}}dG(\mathbf{x}, e)$,
$dQ(\mathbf{x}, e) \equiv ((1-\theta)1\!\!1_{\{\theta\geq\theta_0\}} + (1-\theta_0)1\!\!1_{\{\theta<\theta_0\}})dG(\mathbf{x}, e) + (\theta_0-\theta)1\!\!1_{\{\theta<\theta_0\}}dF(\mathbf{x}, e)$,
$dR(\mathbf{x}, e) \equiv 1\!\!1_{\{\theta<\theta_0\}}dF(\mathbf{x}, e) + 1\!\!1_{\{\theta>\theta_0\}}dG(\mathbf{x}, e)$.

Since under $H_0$, we proved that instead of EL statistic (11) we can consider statistic (25), let us define the following statistic

$$\Lambda_{nk}(\theta_{nk}) = T_{nk}(\theta_{nk}, \tilde{\boldsymbol{\lambda}}(\theta_{nk}), \tilde{\boldsymbol{\beta}}(\theta_{nk}))/(2n), \Lambda_n(0) = \Lambda_n(1) = 0, \quad (29)$$

with $T_{nk}$ given by relation (25), and $\tilde{\boldsymbol{\lambda}}(\theta_{nk}), \tilde{\boldsymbol{\beta}}(\theta_{nk})$ solutions of the system

$$\begin{cases} \frac{\partial T_{nk}(\theta_{nk},\boldsymbol{\lambda},\boldsymbol{\beta})}{2\partial\boldsymbol{\lambda}} = \sum_{i\in I}\frac{\mathbf{g}_i(\boldsymbol{\beta})}{\theta_{nk}+\boldsymbol{\lambda}^t\mathbf{g}_i(\boldsymbol{\beta})} - \sum_{j\in J}\frac{\mathbf{g}_j(\boldsymbol{\beta})}{1-\theta_{nk}-\boldsymbol{\lambda}^t\mathbf{g}_j(\boldsymbol{\beta})} = \mathbf{0}_d, \\ \frac{\partial T_{nk}(\theta_{nk},\boldsymbol{\lambda},\boldsymbol{\beta})}{2\partial\boldsymbol{\beta}} = \sum_{i\in I}\frac{\dot{\mathbf{g}}_i(\boldsymbol{\beta})\boldsymbol{\lambda}}{\theta_{nk}+\boldsymbol{\lambda}^t\mathbf{g}_i(\boldsymbol{\beta})} - \sum_{j\in J}\frac{\dot{\mathbf{g}}_j(\boldsymbol{\beta})\boldsymbol{\lambda}}{1-\theta_{nk}-\boldsymbol{\lambda}^t\mathbf{g}_j(\boldsymbol{\beta})} = \mathbf{0}_d. \end{cases} \quad (30)$$

For any $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$, let the function $K : \Upsilon \times I\!R \times (0,1)$ defined by

$$K(\mathbf{x}, e, \theta) = \theta + \boldsymbol{\lambda}^t \dot{\mathbf{f}}(\mathbf{x}, \boldsymbol{\beta})[e - f(\mathbf{x}, \boldsymbol{\beta}) + f(\mathbf{x}, \boldsymbol{\beta}^{\mathbf{0}})].$$

Let also $\psi(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta}) = \int_\Upsilon \Big( \int_{I\!R} \log K(\mathbf{x}, e, \theta) dP(\mathbf{x}, e) + \int_{I\!R} \log(1 - K(\mathbf{x}, e, \theta)) dQ(\mathbf{x}, e) \Big) dH(\mathbf{x})$ $-\theta \log \theta - (1-\theta)\log(1-\theta)$. We will prove by Theorem 4 that $\psi$ is the limit process of $\Lambda_{nk}$, under $H_1$. Then consider, for a fixed $\theta \in (0,1)$, $\tilde{\tilde{\boldsymbol{\lambda}}}(\theta), \tilde{\tilde{\boldsymbol{\beta}}}(\theta)$ the solutions to the following score equations

$$\begin{cases} \mathbf{z}_1(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta}) = \int_\Upsilon \Big( \int_{I\!R} \frac{\mathbf{g}(\mathbf{x},\boldsymbol{\beta})}{K(\mathbf{x},e,\theta)} dP(\mathbf{x}, e) - \int_{I\!R} \frac{\mathbf{g}(\mathbf{x},\boldsymbol{\beta})}{1-K(\mathbf{x},e,\theta)} dQ(\mathbf{x}, e) \Big) dH(\mathbf{x}) = \mathbf{0}_d, \\ \mathbf{z}_2(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta}) = \int_\Upsilon \Big( \int_{I\!R} \frac{\dot{\mathbf{g}}(\mathbf{x},\boldsymbol{\beta})\boldsymbol{\lambda}}{K(\mathbf{x},e,\theta)} dP(\mathbf{x}, e) - \int_{I\!R} \frac{\dot{\mathbf{g}}(\mathbf{x},\boldsymbol{\beta})}{1-K(\mathbf{x},e,\theta)} dQ(\mathbf{x}, e) \Big) dH(\mathbf{x}) = \mathbf{0}_d, \end{cases} \tag{31}$$

where, $\mathbf{z}_1(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta}) = \partial\psi(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta})/\partial\boldsymbol{\lambda}$ and $\mathbf{z}_2(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta}) = \partial\psi(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta})/\partial\boldsymbol{\beta}$.

We require the following assumptions for the next theorems :

**(A5)** The matrix $\begin{pmatrix} -\frac{\partial \mathbf{z}_1(\theta,\boldsymbol{\lambda},\boldsymbol{\beta})}{\partial \boldsymbol{\lambda}} & -\frac{\partial \mathbf{z}_1(\theta,\boldsymbol{\lambda},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \\ -\frac{\partial \mathbf{z}_2(\theta,\boldsymbol{\lambda},\boldsymbol{\beta})}{\partial \boldsymbol{\lambda}} & -\frac{\partial \mathbf{z}_2(\theta,\boldsymbol{\lambda},\boldsymbol{\beta})}{\partial \beta} \end{pmatrix}$ is nonsingular for all $\theta \in (0,1)$.

**(A6)**

$$\int_\Upsilon \int_{I\!R} \Big( \frac{\mathbf{g}(\mathbf{x},\boldsymbol{\beta})\mathbf{g}^t(\mathbf{x},\boldsymbol{\beta})}{K^2} + \frac{\mathbf{g}(\mathbf{x},\boldsymbol{\beta})\mathbf{g}^t(\mathbf{x},\boldsymbol{\beta})}{(1-K)^2} \Big) d(F(\mathbf{x}, e) + G(\mathbf{x}, e)) < \infty,$$

$$\int_\Upsilon \int_{I\!R} \Big( \frac{\dot{\mathbf{g}}(\mathbf{x},\boldsymbol{\beta})\dot{\mathbf{g}}(\mathbf{x},\boldsymbol{\beta})}{K^2} + \frac{\dot{\mathbf{g}}(\mathbf{x},\boldsymbol{\beta})\dot{\mathbf{g}}(\mathbf{x},\boldsymbol{\beta})}{(1-K)^2} \Big) d(F(\mathbf{x}, e) + G(\mathbf{x}, e)) < \infty, \quad \text{and}$$

$$\int_\Upsilon \int_{I\!R} \Big( \frac{\ddot{\mathbf{g}}(\mathbf{x},\boldsymbol{\beta})}{K^2} + \frac{\ddot{\mathbf{g}}(\mathbf{x},\boldsymbol{\beta})}{(1-K)^2} \Big) d(F(\mathbf{x}, e) + G(\mathbf{x}, e)) < \infty.$$

**(A7)** The functions $f(\mathbf{x}, \boldsymbol{\beta})$ and $\dot{\mathbf{f}}(\mathbf{x}, \boldsymbol{\beta})$ are equicontinuous in $\boldsymbol{\beta}$ on $\Gamma$.

**Remark 3** *A sufficient condition for the equicontinuity of the functions $f(\mathbf{x}, \boldsymbol{\beta})$ and $\dot{\mathbf{f}}(\mathbf{x}, \boldsymbol{\beta})$ is that they are Lipschitzian with respect to $\boldsymbol{\beta}$ on $\Gamma$.*

Following theorem shows that if $\theta_{nk}$ converges to the true value $\theta_0$, then the maximum EL test statistic converges to the maximum of its limit distribution.

**Theorem 4** *Under the alternative hypothesis $H_1$, if the assumptions (A1)-(A7) are satisfied and $\lim_{n\to\infty} \theta_{nk} = \theta \in (0,1)$ then $\Lambda_{nk}(\theta_{nk}) \xrightarrow[n\to\infty]{a.s.} \psi(\theta, \tilde{\tilde{\boldsymbol{\lambda}}}(\theta), \tilde{\tilde{\boldsymbol{\beta}}}(\theta))$, where $\psi(\theta, \tilde{\tilde{\boldsymbol{\lambda}}}(\theta), \tilde{\tilde{\boldsymbol{\beta}}}(\theta))$ is a strictly increasing function on $(0, \theta_0)$ decreasing on $(\theta_0, 1)$ and $\max_{0\leq\theta\leq1} \psi(\theta, \tilde{\tilde{\boldsymbol{\lambda}}}(\theta), \tilde{\tilde{\boldsymbol{\beta}}}(\theta)) = \psi(\theta_0, \tilde{\tilde{\boldsymbol{\lambda}}}(\theta_0), \tilde{\tilde{\boldsymbol{\beta}}}(\theta_0))$.*

*Proof.* We will prove this theorem in three steps.
*Step 1.* We first prove that, for all fixed $\theta \in (0,1)$, we have

$$\arg\max_{(\boldsymbol{\lambda},\boldsymbol{\beta})} T_{nk}(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta}) \xrightarrow[n\to\infty]{a.s.} \arg\max_{(\boldsymbol{\lambda},\boldsymbol{\beta})} \psi(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta}). \tag{32}$$

Obviously, by the law of large numbers, for all $(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta}) \in (0,1) \times I\!\!R \times \Gamma$, we have $(2n)^{-1} T_{nk}(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta}) \overset{a.s.}{\underset{n \to \infty}{\longrightarrow}} \psi(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta})$. On the other hand, by the assumption (A5), $\arg\max_{(\boldsymbol{\lambda}, \boldsymbol{\beta})} \psi(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta})$ is the unique solution of the system (31). Seen the assumptions (A6) and (A7), the function $(2n)^{-1} T_{nk}(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta})$ is equicontinuous and bounded in $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$. Then, using Theorem 1.12.1 of [20], we have that the convergence of $(2n)^{-1} T_{nk}(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta})$ to $\psi(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta})$ is uniform in $(\boldsymbol{\lambda}, \boldsymbol{\beta})$. Taking into account that the solution of system (30) is unique, we obtain relation (32).
*Step 2.* We show that

$$\max_{\theta_{nk}} \Lambda_{nk}(\theta_{nk}) \overset{a.s.}{\underset{n \to \infty}{\longrightarrow}} \max_{\theta} \psi(\theta, \tilde{\tilde{\boldsymbol{\lambda}}}(\theta), \tilde{\tilde{\boldsymbol{\beta}}}(\theta)), \tag{33}$$

with $\tilde{\tilde{\boldsymbol{\lambda}}}(\theta)$ and $\tilde{\tilde{\boldsymbol{\beta}}}(\theta)$ the solutions of score equations (31). By similar calculations as in the proof of Theorem 2, taking into account the Step 1, we can show that, for $\theta = \lim_{n \to \infty} \theta_{nk}$, we have

$$\begin{aligned}\Lambda_{nk}(\theta_{nk}) = & \frac{1}{n} \sum_{i \in I} \log\left[\theta + \tilde{\tilde{\boldsymbol{\lambda}}}^t(\theta) \mathbf{g}_i(\tilde{\tilde{\boldsymbol{\beta}}}(\theta))\right] + \frac{1}{n} \sum_{j \in J} \log\left[(1-\theta) - \tilde{\tilde{\boldsymbol{\lambda}}}^t(\theta) \mathbf{g}_j(\tilde{\tilde{\boldsymbol{\beta}}}(\theta))\right] \\ & - \theta \log\theta - (1-\theta)\log(1-\theta) + o_{I\!\!P}(1).\end{aligned}$$

The above equation, together with the law of large numbers, imply that $\Lambda_{nk}(\theta_{nk}) \overset{a.s.}{\underset{n \to \infty}{\longrightarrow}} \psi(\theta, \tilde{\tilde{\boldsymbol{\lambda}}}(\theta), \tilde{\tilde{\boldsymbol{\beta}}}(\theta))$, where $\theta = \lim_{n \to \infty} \theta_{nk}$.

For $\theta \notin \{0, 1, \theta_0\}$, partial derivative $\partial \psi(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta}) / \partial \theta$ becomes

$$\begin{aligned}&\int_\Upsilon \int_{I\!\!R} \Big[ [\log K(\mathbf{x}, e, \theta) 1\!\!1_{\{\theta < \theta_0\}} dF(\mathbf{x}, e) + \log K(\mathbf{x}, e, \theta) 1\!\!1_{\{\theta > \theta_0\}} dG(\mathbf{x}, e)] \\ &- [\log(1 - K(\mathbf{x}, e, \theta)) 1\!\!1_{\{\theta < \theta_0\}} dF(\mathbf{x}, e) + \log(1 - K(\mathbf{x}, e, \theta)) 1\!\!1_{\{\theta > \theta_0\}} dG(\mathbf{x}, e)] \Big] dH(\mathbf{x}) \\ &+ \log(1-\theta) - \log\theta.\end{aligned}$$

On the other hand, we have that, $dR(\mathbf{x}, e) = 1\!\!1_{\{\theta < \theta_0\}} dF(\mathbf{x}, e) + 1\!\!1_{\{\theta > \theta_0\}} dG(\mathbf{x}, e)$.
Hence,

$$\begin{aligned}\frac{\partial \psi(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta})}{\partial \theta} = & \int_\Upsilon \int_{I\!\!R} [\log K(\mathbf{x}, e, \theta) - \log(1 - K(\mathbf{x}, e, \theta))] dR(\mathbf{x}, e) dH(\mathbf{x}) \\ & - \log\theta + \log(1-\theta).\end{aligned}$$

Because $\tilde{\tilde{\boldsymbol{\lambda}}}^t(\theta) \mathbf{g}(\mathbf{x}, \tilde{\tilde{\boldsymbol{\beta}}}(\theta)) = K(\mathbf{x}, e, \theta) - \theta$ and $\tilde{\tilde{\boldsymbol{\lambda}}}^t(\theta) \mathbf{z}_1(\theta, \tilde{\tilde{\boldsymbol{\lambda}}}(\theta), \tilde{\tilde{\boldsymbol{\beta}}}(\theta)) = 0$, we obtain

$$\int_\Upsilon \big[ \int_{I\!\!R} (1 - \frac{\theta}{K(\mathbf{x}, e, \theta)}) dP(\mathbf{x}, e) + \int_{I\!\!R} (1 - \frac{1-\theta}{1 - K(\mathbf{x}, e, \theta)}) dQ(\mathbf{x}, e) \big] dH(\mathbf{x}) = 0. \tag{34}$$

On the other hand, we have $\mathbf{z}_2(\theta, \tilde{\tilde{\boldsymbol{\lambda}}}(\theta), \tilde{\tilde{\boldsymbol{\beta}}}(\theta)) = \mathbf{0}_d$. Then

$$\int_\Upsilon \big[ \dot{\mathbf{g}}(\mathbf{x}, \tilde{\tilde{\boldsymbol{\beta}}}(\theta)) \int_{I\!\!R} \frac{dP(\mathbf{x}, e)}{K(\mathbf{x}, e, \theta)} \big] dH(\mathbf{x}) = \int_\Upsilon \big[ \dot{\mathbf{g}}(\mathbf{x}, \tilde{\tilde{\boldsymbol{\beta}}}(\theta)) \int_{I\!\!R} \frac{dQ(\mathbf{x}, e)}{1 - K(\mathbf{x}, e, \theta)} \big] dH(\mathbf{x}) = 0.$$

Since $\int_\Upsilon \big[\int_{I\!R} dP(\mathbf{x},e)+\int_{I\!R} dQ(\mathbf{x},e)\big]dH(\mathbf{x})=1$, relation (34) becomes

$$1-\theta\int_\Upsilon\int_{I\!R}\Big[\frac{dP(\mathbf{x},e)}{K(\mathbf{x},e,\theta)}-\frac{dQ(\mathbf{x},e)}{1-K(\mathbf{x},e,\theta)}\Big]dH(\mathbf{x})-\int_\Upsilon\int_{I\!R}\frac{dQ(\mathbf{x},e)}{1-K(\mathbf{x},e,\theta)}dH(\mathbf{x})=0. \tag{35}$$

This relation is true for all $\theta\in(0,1)$. If we take $\theta=0$ and afterward $\theta=1$, relation (35) implies

$$\int_\Upsilon\int_{I\!R}\frac{dP(\mathbf{x},e)}{K(\mathbf{x},e,\theta)}dH(\mathbf{x})=\int_\Upsilon\int_{I\!R}\frac{dQ(\mathbf{x},e)}{1-K(\mathbf{x},e,\theta)}dH(\mathbf{x})=1. \tag{36}$$

The relation (33) is proved in a similar way as the proof of Theorem 3.2 of [14], using relations (35) and (37).
*Step 3.* Similar as in the proof of Theorem 3.2 of [14], we can prove that for all $\gamma\in(0,\min(\theta_0,1-\theta_0)$, we have

$$\max_{|k-n\theta_0|\geq n\gamma}\Lambda_{nk}(\theta_{nk})\xrightarrow[n\to\infty]{a.s.}\max_{|\theta-\theta_0|\geq\gamma}\psi(\theta,\tilde{\tilde{\boldsymbol{\lambda}}}(\theta),\tilde{\tilde{\boldsymbol{\beta}}}(\theta)). \tag{37}$$

Which implies $\lim_{n\to\infty}I\!P[|\arg\max_k\Lambda_{nk}(\theta_{nk})-\theta_0|\geq\gamma]=0$. $\square$

**Corollary 1** *The proof of Theorem 4 implies that maximum EL estimator of $\theta_0$ defined by $\tilde{\theta}_n\equiv\min\{\tilde{\theta}_{nk};\tilde{\theta}_{nk}=\arg\max_{\theta_{nk}\in[\Theta_{1n},\Theta_{2n}]}T_{nk}(\theta_{nk},\hat{\boldsymbol{\lambda}}(\theta_{nk}),\hat{\boldsymbol{\beta}}(\theta_{nk}))\}$ satisfies the property that $\tilde{\theta}_n-\theta_{n0}\xrightarrow[n\to\infty]{I\!P}0$. Taking into account Remark 2, we have also $\hat{\theta}_n-\theta-n0\to0$ in probability.*

We prove by the following theorem that statistical test $\tilde{T}_n$ given by (26) has the asymptotic power 1.

**Theorem 5** *Under assumptions (A1)-(A7), the power of the empirical likelihood ratio test $\tilde{T}_n$ converges to 1.*

*Proof.* By Theorem 3 and relation (25), in order to study the test power, we consider the probability

$$1-I\!P\big[A(\log u(n))\tilde{T}_n^{\frac{1}{2}}\leq t+D(\log u(n))\big], \tag{38}$$

where $(\hat{\boldsymbol{\lambda}}(\theta_{nk}),\hat{\boldsymbol{\beta}}(\theta_{nk}))$ are the solutions of score equations (12) and (13).
By Theorem 4 we have shown that, under $H_1$, for $\lim_{n\to\infty}\theta_{nk}=\theta\in(0,1)$, we have

$$\Lambda_{nk}(\theta_{nk})\xrightarrow[n\to\infty]{a.s.}\psi(\theta,\tilde{\tilde{\boldsymbol{\lambda}}}(\theta),\tilde{\tilde{\boldsymbol{\beta}}}(\theta)).$$

Let us denote by $v_n$ the convergence rate of $\Lambda_{nk}(\theta_{nk})$ to 0. By elementary calculations, we obtain that

$$\lim_{n\to\infty}I\!P\Big[\Big(\frac{\log\log u(n)}{v_n}\Big)^{1/2}\Big(\frac{\tilde{T}_n}{2n}\Big)^{1/2}\leq\frac{t}{\sqrt{2nv_n}}+\frac{D(\log u(n))}{\sqrt{2nv_n}}\Big].$$

Since $v_n$ is the convergence rate of $\Lambda_{nk}(\theta)$, we have $n^{-1}\tilde{T}_n = O_{I\!P}(v_n)$. Then for all $\epsilon > 0$, there exists $\varrho > 0$ such that $I\!P[2^{-1}n^{-1}(v_n)^{-1}|\tilde{T}_n| \geq \varrho] \leq \epsilon$. Which implies that $(v_n)^{-1/2}(2n)^{-1/2}\tilde{T}_n^{1/2}$ is bounded with a probability close to 1. Hence, $(\log\log u(n)/v_n)^{1/2}(\tilde{T}_n/2n)^{1/2}$ is not bounded with a probability close to 1. With this results, considering $t = (2nv_n)^{1/2}$ in relation (38), we obtain

$$\lim_{n\to\infty} I\!P\Big[\Big(\frac{\log\log u(n)}{v_n}\Big)^{1/2}\Big(\frac{\tilde{T}_n}{2n}\Big)^{1/2} \leq \frac{D(\log u(n))}{\sqrt{2nv_n}}\Big] = 0.$$

The theorem follows. □

We emphasise that, similar results to Theorems 3, 4 and 5 were obtained for simpler models : test to detecting a change in distribution sequence [24], or a change in the parameters of a linear model [14].

## 3 Two change-points test

In this section, we consider the epidemic model. We assume under alternative hypothesis, denoted $H_2$, that the model have two change-points $k_1$ and $k_2$ $(1 < k_1 < k_2 < n)$, such that the model of the first and the third segment is the same. More specifically, the regression model can be written

$$H_2 : \mathbf{Y}_i = \begin{cases} f(\mathbf{X}_i, \boldsymbol{\beta}_1) + \varepsilon_i & i = 1, \cdots, k_1 \\ f(\mathbf{X}_i, \boldsymbol{\beta}_2) + \varepsilon_i & i = k_1 + 1, \cdots, k_2 \\ f(\mathbf{X}_i, \boldsymbol{\beta}_1) + \varepsilon_i & i = k_2 + 1, \cdots, n. \end{cases} \tag{39}$$

Therefore, we want to test the null hypothesis $H_0$ of no-change, against the alternative hypothesis $H_2$.

Under the hypothesis $H_2$, we consider the following two sets, $I' = \{1, ..., k_1, k_2 + 1, ..., n\}$ and $J' = \{k_1 + 1, ..., k_2\}$, and we define the corresponding probability vectors $(u_1 \cdots, u_{k_1}, u_{k_2+1}, \cdots, u_n)$ and $(v_{k_1+1}, \cdots, v_{k_2})$, where $u_i \equiv P[Y_i = y_i]$, and $v_j \equiv P[Y_j = y_j]$ denotes the probability to observe the value $y_i$ (respectively $y_j$), for the dependent variable $Y_i$ (respectively $Y_j$), for $i \in I'$ and $j \in J'$. Obviously, $\sum_{i\in I'} u_i = 1$ and $\sum_{j\in J'} v_j = 1$.

Under hypothesis $H_0$, the profile EL ratio for $\boldsymbol{\beta}$ is

$$\begin{aligned} U_0'(\boldsymbol{\beta}) = \sup_{(u_1,\cdots,u_{k_1},u_{k_2+1},\cdots,u_n)} \sup_{(v_{k_1+1},\cdots,v_{k_2})} \Big\{ \prod_{i\in I'}(n - k_2 + k_1)u_i \prod_{j\in J'}(k_2 - k_1)v_j; \\ \sum_{i\in I'} u_i = \sum_{j\in J'} v_j = 1 \sum_{i\in I'} u_i \mathbf{g}_i(\boldsymbol{\beta}) = \sum_{j\in J'} v_i \mathbf{g}_j(\boldsymbol{\beta}) = \mathbf{0}_d \Big\}. \end{aligned}$$

Under hypothesis $H_2$, the profile EL ratio for $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ has the form

$$\begin{aligned} U_1'(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \sup_{(u_1,\cdots,u_{k_1},u_{k_2+1},\cdots,u_n)} \sup_{(v_{k_1+1},\cdots,v_{k_2})} \Big\{ \prod_{i\in I'}(n - k_2 + k_1)u_i \prod_{j\in J'}(k_2 - k_1)v_j; \\ \sum_{i\in I'} u_i \mathbf{g}_i(\boldsymbol{\beta}_1) = \sum_{j\in J'} v_j \mathbf{g}_j(\boldsymbol{\beta}_2) = \mathbf{0}_d \Big\}. \end{aligned}$$

Then, in order to test $H_0$ against $H_2$, we consider the profile EL ratio $U'_0(\boldsymbol{\beta})/U'_1(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)$.

Similarly as in Section 2, when we tested a single change-point, using Lagrange multipliers, we obtain that under hypothesis $H_0$,the probabilities $u_i$, $v_j$ are

$$u_i = \frac{1}{(n-k_2+k_1)+n\boldsymbol{\lambda}_1^t\mathbf{g}_i(\boldsymbol{\beta})} \qquad v_j = \frac{1}{(k_2-k_1)-n\boldsymbol{\lambda}_2^t\mathbf{g}_j(\boldsymbol{\beta})}. \tag{40}$$

Using the similar arguments as in the proof of Theorem 1, we deduce that the asymptotic distribution of $-2\log U'_1(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)$ is $\chi^2(3d)$ and then we can consider the test statistic $-2\log U'_0(\boldsymbol{\beta})$. We restricted to the case where $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ satisfy the constraint $\tilde{\mathbf{V}}_1(\boldsymbol{\beta})\boldsymbol{\lambda}_1 = \tilde{\mathbf{V}}_2(\boldsymbol{\beta})\boldsymbol{\lambda}_2$, with $\tilde{\mathbf{V}}_1(\boldsymbol{\beta}) = (n+k_1-k_2)^{-1}\sum_{i\in I'}\dot{\mathbf{g}}(\boldsymbol{\beta})$ and $\tilde{\mathbf{V}}_2(\boldsymbol{\beta}) = (k_2-k_1)^{-1}\sum_{j\in J'}\dot{\mathbf{g}}(\boldsymbol{\beta})$. In this case, considering the parameter $\theta_{n,k_1,k_2} = n^{-1}(n-k_2+k_1)$, that depends on two change-points $k_1$, $k_2$, we will consider the test statistic

$$2\sum_{i\in I'}\log\left[1+\frac{1}{\theta_{n,k_1,k_2}}\boldsymbol{\lambda}^t\mathbf{g}_i(\boldsymbol{\beta})\right]+2\sum_{j\in J'}\log\left[1-\frac{1}{1-\theta_{n,k_1,k_2}}\boldsymbol{\lambda}^t\tilde{\mathbf{V}}_1(\boldsymbol{\beta})\tilde{\mathbf{V}}_2^{-1}(\boldsymbol{\beta})\mathbf{g}_j(\boldsymbol{\beta})\right]. \tag{41}$$

and $\hat{\boldsymbol{\lambda}}(\theta_{n,k_1,k_2})$, $\hat{\boldsymbol{\beta}}(\theta_{n,k_1,k_2})$ solutions of the score equations of this random process equal to zero. We can show, as in Section 2, that statistic (41) is, under hypothesis $H_0$, asymptotically equivalent to the statistic

$$U(\theta_{n,k_1,k_2},\boldsymbol{\lambda},\boldsymbol{\beta}) \equiv 2\sum_{i\in I'}\log\left[1+\theta^{-1}_{n,k_1,k_2}\boldsymbol{\lambda}^t\mathbf{g}_i(\boldsymbol{\beta})\right]+2\sum_{j\in J'}\log\left[1-(1-\theta_{n,k_1,k_2})^{-1}\boldsymbol{\lambda}^t\mathbf{g}_j(\boldsymbol{\beta})\right].$$

Then, we will consider for testing null hypothesis $H_0$ against $H_2$ the test statistic $\max_{1<k_1<k_2<n}\{U(\theta_{n,k_1,k_2},\hat{\boldsymbol{\lambda}}(\theta_{n,k_1,k_2}),\hat{\boldsymbol{\beta}}(\theta_{n,k_1,k_2}))\}$.

In the case when $k_1$ or $k_2-k_1$ have a small value, the maximum empirical likelihood may not exist. In this case, the proposed test may not detect the presence of change in the model. For the empirical likelihood maximum always exists, we consider two natural numbers $\Theta_{n1}$ and $\Theta_{n2}$, such as $\Theta_{n1} < k_1 < k_2 < n-\Theta_{n2}$. Finally, the test statistic for testing $H_0$ against $H_2$ becomes

$$\max_{\Theta_{n1}<k_1<k_2<n-\Theta_{n2}}\{U(\theta_{n,k_1,k_2},\hat{\boldsymbol{\lambda}}(\theta_{n,k_1,k_2}),\hat{\boldsymbol{\beta}}(\theta_{n,k_1,k_2}))\}.$$

We easily obtain the corresponding statistic given in Remark 2 by relation (28) to facilitate the practical utilization of the test statistic.

## 4 Simulation study

In this section, we report a simulation study by Monte Carlo method, in order to evaluate the performance of the proposed test statistics. Firstly, when the nonlinear regression model have a single change-point, secondly, when this same model have two change-points.

All simulations were performed using the R language. The program codes are

**Table 1** Critical values $c_\alpha$, for $\alpha$=0.05. Empirical asymptotic power on 500 Monte Carlo replications, when $\varepsilon \sim \mathcal{N}(0,1)$.

| n | $k_0$ | $c_\alpha$ | power |
|---|---|---|---|
| 1000 | 600 | 1.544 | 1 |
| 800 | 500 | 1.492 | 1 |
| 600 | 400 | 1.434 | 1 |
| 400 | 250 | 1.340 | 1 |
| 200 | 75 | 1.133 | 1 |

available from the authors.
We consider the nonlinear function

$$f(x, \boldsymbol{\beta}) = a\frac{1-x^b}{b}. \tag{42}$$

with $\boldsymbol{\beta} = (a, b) \in [-100, 100] \times [0.1, 20]$.

## 4.1 Model with a single change-point

For the nonlinear function of (42), the following two-phase (one change-point) nonlinear model is considered under $H_1$

$$Y_i = a_1 \frac{1-X_i^{b_1}}{b_1} 1\!\!1_{i \leq k_0} + a_2 \frac{1-X_i^{b_2}}{b_2} 1\!\!1_{i > k_0} + \varepsilon_i, \qquad i = 1, \cdots, n \tag{43}$$

with $X_i = i/1000$, $n = 1000$ and true value of parameters $a_1^0 = 10$, $b_1^0 = 2$, $a_2^0 = 7$, $b_2^0 = 1.75$. Under hypothesis $H_0$, the true parameters are $a^0 = 10$, $b^0 = 2$. The same model was considered in [5], where the model was estimated by the penalized least absolute deviation method.
The change absence against one-change in model is tested using the (approached) maximum empirical likelihood statistic $T(\theta_{nk})$ given by (28).
In order to calculate the empirical test size, an without change-point model is considered and we count, the number of times, on the Monte Carlo replications when we obtain $\left(\max_{\theta_{nk}} T(\theta_{nk})\right)^{1/2} \geq c_\alpha$. For a fixed size $\alpha \in (0, 1)$, critical value $c_\alpha$ is calculated in accordance with Corollary 1 :

$$c_\alpha = \frac{-\log(-\log \alpha) + D(\log u(n))}{A(\log u(n))}.$$

For theoretical size $\alpha = 0.05$, we first calculate critical values $c_\alpha$, varying the sample size $n$ from 200 to 1000 (see Table 1).
For model (43) with Gaussian standardized errors, 500 Monte Carlo replications were performed. We also present in Table 1 the empirical asymptotic power, using statistic test (28) for different position of change-point. For any change-point location, the asymptotic test power is 1.

**Table 2** Empirical size for four error distributions on 500 Monte Carlo replications, n=1000, $\alpha$=0.05.

| | Normal | Exponential | $\chi^2$ | Student |
|---|---|---|---|---|
| Empirical size | 0 | 0 | 0 | 0.08 |

We fix sample size $n = 1000$, theoretical test size $\alpha = 0.05$ and we vary error distribution. In order to calculate the empirical size of test (type I error probabilities), 500 Monte Carlo replications are realized for different error distributions: $\varepsilon_i = \mathcal{N}(0,1)$, $\varepsilon_i = 2\mathcal{E}xp(2) - 1$, $\varepsilon_i = 1/\sqrt{6}(\chi^2(3) - 3)$ and $\varepsilon_i = 2/\sqrt{6}t(6)$, where $\mathcal{N}(0,1)$, $\mathcal{E}xp(2)$, $\chi^2(3)$ and $t(6)$ are standard normal distribution, exponential distribution with mean 1/2, chi-square distribution with degree of freedom 3 and Student distribution with degree of freedom 6, respectively. In all cases, excepting for Student distribution (when the empirical size is slightly larger than 0.05), the empirical size is 0 (see Tables 2). For the same four error distributions, but for model with a change-point in $k_0$, by 500 Monte Carlo model replications, for different change-point location: $k_0 \in \{200, 400, 600, 800\}$, we obtain that the empirical power is 1, in any case.

As mentioned in Remark 2, one can also estimate the change-point location by EL method. Table 3 shows descriptive statistics: minimal, maximal values, mean and median of $\hat{k}_n$ for 500 Monte Carlo replications. In view of the results presented in Table 3, for different error distributions and for different positions of the change in the measurement interval, we deduce that the proposed estimation method approaches very well the true value $k^0$, regardless of the error distribution and of the change-point position on the interval $[1 : n]$. The results are not influenced by error distribution, then outlier presence has no effect on the estimate. Nevertheless, precision is influenced when the change is in the right part of the measurement interval. Note that, in all situations the median and the mean of the change-point estimations coincide or is very close to the true value.

## 4.2 Model with two change-points

For nonlinear function of (42), under hypothesis $H_2$, we consider the following three-phase (two change-points) model

$$Y_i = a_1 \frac{1 - X_i^{b_1}}{b_1} 1\!\!1_{i \leq k_1} + a_2 \frac{1 - X_i^{b_2}}{b_2} 1\!\!1_{k_1 < i \leq k_2} + a_1 \frac{1 - X_i^{b_1}}{b_1} 1\!\!1_{k_2 < i \leq n} + \varepsilon_i, \quad (44)$$

with $X_i = i/1000$, $n = 1500$ and the true value of parameters $a_1^0 = 10$, $b_1^0 = 2$, $a_2^0 = 7$, $b_2^0 = 1.75$. Under null hypothesis $H_0$ the true parameters are $a^0 = 10$, $b^0 = 2$.

In Table 4 we give results after 150 Monte Carlo replications in order to calculate the empirical power of test, for $n = 1500$. We deduce that empirical size is zero and asymptotic test power is 1.

**Table 3** Descriptive statistics for $\hat{k}_n$ by EL method( model with two phases). $n = 1000$, 500 Monte Carlo replications.

| error distribution | $k^0$ | $\hat{k}_n$ | | | |
|---|---|---|---|---|---|
| | | min($\hat{k}_n$) | max($\hat{k}_n$) | mean($\hat{k}_n$) | median($\hat{k}_n$) |
| $\varepsilon_i \sim \mathcal{N}(0,1)$ | 200 | 168 | 211 | 198 | 200 |
| | 400 | 391 | 425 | 401 | 400 |
| | 600 | 552 | 616 | 598 | 600 |
| | 800 | 701 | 823 | 782 | 794 |
| $\varepsilon_i \sim 2/\sqrt{6}t(6)$ | 200 | 170 | 222 | 196 | 199 |
| | 400 | 385 | 420 | 400 | 400 |
| | 600 | 594 | 600 | 598 | 599 |
| | 800 | 706 | 819 | 776 | 789 |
| $\varepsilon_i \sim 2\mathcal{E}xp(2) - 1$ | 200 | 181 | 211 | 199 | 199 |
| | 400 | 388 | 420 | 401 | 400 |
| | 600 | 565 | 616 | 599 | 600 |
| | 800 | 702 | 820 | 773 | 788 |
| $\varepsilon_i \sim 1/\sqrt{6}(\chi^2(3) - 3)$ | 200 | 153 | 224 | 198 | 200 |
| | 400 | 390 | 422 | 401 | 400 |
| | 600 | 557 | 621 | 598 | 600 |
| | 800 | 700 | 820 | 785 | 795 |

**Table 4** Test with two change-points.

| $k_1$ | $k_2$ | power |
|---|---|---|
| no-change | | 0 |
| 100 | 900 | 1 |
| 200 | 500 | 1 |
| 400 | 600 | 1 |
| 600 | 900 | 1 |

### 4.3 Comparison with LS test

On data considered in sub-section 4.1 for $\varepsilon \sim \mathcal{N}(0,1)$ and $n = 1000$ we apply the method proposed by [3], where the estimation method and the associated test is least squares. This study is realized by computing the test statistic $\sup F(0:1)$ given in [3]. Under hypothesis $H_1$ that the model has a change-point in $k^0 = 600$, 500 Monte Carlo simulations each time give that the test statistic value exceeds critical value of 12.85 ( see [2]), then the null hypothesis $H_0$ is rejected and hence the power test is 1. Whereas if we generate the values $Y_i$ without change-point for gaussian errors, then, the test statistic value always exceeds critical. Hence the empirical size of the test proposed by [3] is 1, a result significantly worse than that obtained by our test. We note that (see Table 5) if under $H_1$, the true change-point is off-centred in the measurement interval, because of the function nonlinearity, numerical problem arise for the LS estimation method. This is symbolised by "???" in Table 5. The same problem appears when the errors are not gaussian, regardless of the positionn

**Table 5** Descriptive statistics for the change-point estimations by LS method( model with two phases) $n = 1000$, 500 Monte Carlo replications.

| error distribution | $k^0$ | $\hat{k}_n$ | | | |
|---|---|---|---|---|---|
| | | $\min(\hat{k}_n)$ | $\max(\hat{k}_n)$ | $\text{mean}(\hat{k}_n)$ | $\text{median}(\hat{k}_n)$ |
| $\varepsilon_i \sim \mathcal{N}(0,1)$ | 200 | ??? | ??? | ??? | ??? |
| | 400 | 396 | 400 | 399 | 400 |
| | 600 | 595 | 605 | 600 | 600 |
| | 800 | ??? | ??? | ??? | ??? |

of the change-point in the measurement interval. In contrast, we have seen that the EL test works for any error distribution and any change-point position.

## 5 Appendix

The following lemma will be used in the proof of propositions, theorems and of other lemmas.

**Lemma 1** *Let $\mathbf{X} = (X_1, \cdots, X_p)$ a random vector (column), with the random variables $X_1, \cdots, X_p$ not necessarily independent, and $\boldsymbol{M} = (m_{ij})_{1 \leq i,j \leq p}$, such that $\boldsymbol{M} = \mathbf{X}\mathbf{X}^t$. If for j=1, ..., p, we have*

$$for\ all\ \eta_j > 0, there\ exists\ \delta_j > 0 \quad such\ that\ \ I\!\!P[|X_j| \geq \delta_j] \leq \eta_j, \tag{45}$$

*then*
*(i) $I\!\!P\big[\|\mathbf{X}\|_1 \geq p \max_{1 \leq j \leq p} \delta_j\big] \leq \max_{1 \leq j \leq p} \eta_j$,*
*(ii) $I\!\!P\big[\|\mathbf{X}\|_2 \geq \sqrt{p} \max_{1 \leq j \leq p} \delta_j\big] \leq \max_{1 \leq j \leq p} \eta_j$,*
*(iii) $I\!\!P\big[\|\boldsymbol{M}\|_1 \geq p \max_{1 \leq i,j \leq p}\{\delta_i^2, \delta_j^2\}\big] \leq \max_{1 \leq i,j \leq p}\{\eta_i^2, \eta_j^2\}$,*
*where $\|\boldsymbol{M}\|_1 = \max_{1 \leq j \leq p}\{\sum_{i=1}^p |m_{ij}|\}$ is the subordinate norm to the vector norm $\|.\|_1$.*

*Proof of Lemma 1.* $(i)$ Using the relation (45), we can write

$$I\!\!P[\|\mathbf{X}\|_1 \geq p \max_{1 \leq j \leq p} \delta_j] \leq I\!\!P[p \max_{1 \leq j \leq p} |X_j| \geq p \max_{1 \leq j \leq p} \delta_j] \leq \max_{1 \leq j \leq p} \eta_j.$$

$(ii)$ The relation (45) is equivalent to $I\!\!P\big[X_j^2 \geq \delta_j^2\big] \leq \eta_j$, which implies that

$$I\!\!P[\|\mathbf{X}\|_2^2 \geq p \max_{1 \leq j \leq p} \delta_j^2] = I\!\!P[\max_{1 \leq j \leq p} X_j^2 \geq \max_{1 \leq j \leq p} \delta_j^2] \leq \max_{1 \leq j \leq p} \eta_j.$$

$(iii)$ For $1 \leq i, j \leq p$, we have

$I\!\!P[|X_i X_j| \geq \max\{\delta_i^2, \delta_j^2\}] \leq I\!\!P[\max\{X_i^2, X_j^2\} \geq \max\{\delta_i^2, \delta_j^2\}] \leq \max\{\eta_i^2, \eta_j^2\}$.

Then, $I\!\!P[|m_{ij}| \geq \max\{\delta_i^2, \delta_j^2\}] \leq \max\{\eta_i^2, \eta_j^2\}$. Hence, for each $1 \leq j \leq p$,

$$I\!\!P[\sum_{i=1}^p |m_{ij}| \geq p \max\{\delta_i^2, \delta_j^2\}] \leq I\!\!P\big[p \max_{1 \leq i \leq p} |m_{ij}| \geq p \max\{\delta_i^2, \delta_j^2\}\big] \leq \max\{\eta_i^2, \eta_j^2\}.$$

$\square$

**Lemma 2** *Let the $\eta$-neighbourhood of $\boldsymbol{\beta^0}$, $\mathcal{V}_\eta(\boldsymbol{\beta^0}) = \{\boldsymbol{\beta} \in \Gamma; \|\boldsymbol{\beta} - \boldsymbol{\beta^0}\|_2 \leq \eta\}$, with $\eta \to 0$. Then, under assumptions (A1)-(A4), for all $\epsilon > 0$, there exists a positive constant $M > 0$, such that, for all $\boldsymbol{\beta} \in \mathcal{V}_\eta(\boldsymbol{\beta^0})$,*

$$I\!P\big[\|\mathbf{g}_i(\boldsymbol{\beta})\|_1 \geq M\big] \leq \epsilon.$$

*Proof of Lemma 2.* In the following, for simplicity, we denote the functions $\dot{\mathbf{f}}(\mathbf{X}_i, \boldsymbol{\beta})$ by $\dot{\mathbf{f}}_i(\boldsymbol{\beta})$, and $\ddot{\mathbf{f}}(\mathbf{X}_i, \boldsymbol{\beta})$ by $\ddot{\mathbf{f}}_i(\boldsymbol{\beta})$. The Taylor's expansion up the order 2 of $\mathbf{g}_i(\boldsymbol{\beta})$ at $\boldsymbol{\beta} = \boldsymbol{\beta}^0$ is

$$\begin{aligned}\mathbf{g}_i(\boldsymbol{\beta}) = {} & \dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\varepsilon_i + \frac{1}{2}\ddot{\mathbf{f}}_i(\tilde{\boldsymbol{\beta}}_i)(\boldsymbol{\beta} - \boldsymbol{\beta^0})\varepsilon_i - \frac{1}{2}\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})(\boldsymbol{\beta} - \boldsymbol{\beta^0}) \\ & -\frac{1}{6}\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})(\boldsymbol{\beta} - \boldsymbol{\beta^0})^t\ddot{\mathbf{f}}_i(\tilde{\tilde{\boldsymbol{\beta}}}_i)(\boldsymbol{\beta} - \boldsymbol{\beta^0}) - \frac{1}{4}\ddot{\mathbf{f}}_i(\tilde{\boldsymbol{\beta}}_i)(\boldsymbol{\beta} - \boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})(\boldsymbol{\beta} - \boldsymbol{\beta^0}) \\ & -\frac{1}{12}\ddot{\mathbf{f}}_i(\tilde{\boldsymbol{\beta}}_i)(\boldsymbol{\beta} - \boldsymbol{\beta^0})(\boldsymbol{\beta} - \boldsymbol{\beta^0})^t\ddot{\mathbf{f}}_i(\tilde{\tilde{\boldsymbol{\beta}}}_i)(\boldsymbol{\beta} - \boldsymbol{\beta^0}), \end{aligned} \tag{46}$$

where, $\tilde{\boldsymbol{\beta}}_i = \boldsymbol{\beta^0} + \mathbf{u}(\boldsymbol{\beta} - \boldsymbol{\beta^0})$, $\tilde{\tilde{\boldsymbol{\beta}}}_i = \boldsymbol{\beta^0} + \mathbf{v}(\boldsymbol{\beta} - \boldsymbol{\beta^0})$, with $\mathbf{u}, \mathbf{v} \in [0,1]^d$. We note that $\tilde{\boldsymbol{\beta}}_i$ and $\tilde{\tilde{\boldsymbol{\beta}}}_i$ are random vectors which depend on $\mathbf{X}_i$.
For $\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\varepsilon_i$, because $\mathbf{X}_i$ and $\varepsilon_i$ are independent, and $I\!E(\varepsilon_i) = 0$, we have that $I\!E[\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\varepsilon_i] = 0$ and $\mathbb{V}\mathrm{ar}\,[\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\varepsilon_i] = \sigma^2\mathbf{V}$. For the j-th component of $\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})$, by the Bienaymé-Tchebychev's inequality, for $1 \leq j \leq d$, for all $\epsilon_1 > 0$, we have

$$I\!P\big[|\dot{\mathbf{f}}_{ij}(\boldsymbol{\beta^0})\varepsilon_i| \geq \epsilon_1\big] \leq \frac{\sigma^2}{\epsilon_1^2}V_{jj}, \tag{47}$$

where $V_{jj}$ is the j-th term diagonal of the matrix $\mathbf{V}$.
Let $\epsilon > 0$, taking $\epsilon_1 = \sigma\sqrt{6V_{jj}/\epsilon}$ in (47), we obtain $I\!P\big[|\dot{\mathbf{f}}_{ij}(\boldsymbol{\beta^0})\varepsilon_i| \geq \sigma\sqrt{6V_{jj}/\epsilon}\big] \leq \epsilon/6$. Applying Lemma 1 (i), we obtain

$$I\!P\big[\|\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\varepsilon_i\|_1 \geq \frac{\sigma d}{\sqrt{\epsilon}}\max_{1\leq j\leq d}\sqrt{6V_{jj}}\big] \leq \epsilon/6. \tag{48}$$

For the second term of the right-hand side of (46), using assumption (A3), we obtain that for $1 \leq j \leq d$, for all $\epsilon > 0$ there exists $\epsilon_2 > 0$, such that $I\!P\big[|\ddot{\mathbf{f}}_{ij}(\tilde{\boldsymbol{\beta}}_i)| \geq \epsilon_2\big] \leq \epsilon/6$. By Lemma 1 (iii), we have that for all $\epsilon > 0$,

$$I\!P\big[\|\ddot{\mathbf{f}}_i(\tilde{\boldsymbol{\beta}}_i)\|_1 \geq \epsilon_2\big] \leq \frac{\epsilon}{6}. \tag{49}$$

Using Bienaymé-Tchebychev's inequality, and assumption (A1), we obtain that for all $\epsilon > 0$

$$I\!P\big[|\varepsilon_i| > c\big] \leq \frac{\sigma^2}{c}. \tag{50}$$

Recall that $\|\boldsymbol{\beta} - \boldsymbol{\beta^0}\|_2 < \eta$, with $\eta \to 0$. Then, using (49) and (50), we can write that, for all $\epsilon > 0$, there exists $\epsilon_2 > 0$ such that $I\!P\big[\|\ddot{\mathbf{f}}_i(\tilde{\boldsymbol{\beta}}_i)(\boldsymbol{\beta} - \boldsymbol{\beta^0})\varepsilon_i\|_1 \geq \epsilon_2\big] \leq$

$I\!\!P\big[\|\ddot{\mathbf{f}}_i(\tilde{\boldsymbol{\beta}}_i)\|_1|\varepsilon_i|\ \|\boldsymbol{\beta}-\boldsymbol{\beta^0}\|_1 \geq \epsilon_2\big] \leq I\!\!P\big[\|\ddot{\mathbf{f}}_i(\tilde{\boldsymbol{\beta}}_i)\|_1 \geq \epsilon_2/c\eta\big] \leq I\!\!P\big[\|\ddot{\mathbf{f}}_i(\tilde{\boldsymbol{\beta}}_i)\|_1 \geq \epsilon_2\big] \leq \epsilon/6$. Therefore, for all $\epsilon > 0$, there exists $\epsilon_2 > 0$ such that

$$I\!\!P\big[\|\ddot{\mathbf{f}}_i(\tilde{\boldsymbol{\beta}}_i)(\boldsymbol{\beta}-\boldsymbol{\beta^0})\varepsilon_i\|_1 \geq \epsilon_2\big] \leq \frac{\epsilon}{6}. \tag{51}$$

We consider now the term $\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})$ of relation (46). By Markov's inequality, taking also into account assumption (A4), we obtain for $1 \leq j, l \leq d$, for all $\epsilon_3 > 0$, that $I\!\!P\big[|\dot{\mathbf{f}}_{ij}(\boldsymbol{\beta^0})\dot{\mathbf{f}}_{il}^t(\boldsymbol{\beta^0})| \geq \epsilon_3\big] \leq I\!\!E[\dot{\mathbf{f}}_{ij}(\boldsymbol{\beta^0})\dot{\mathbf{f}}_{il}^t(\boldsymbol{\beta^0})]/\epsilon_3$. We choose, for all $\epsilon > 0, \epsilon_3 = 6I\!\!E[\dot{\mathbf{f}}_{ij}(\boldsymbol{\beta^0})\dot{\mathbf{f}}_{il}^t(\boldsymbol{\beta^0})]/\epsilon$. Then, the last relation becomes $I\!\!P\big[|\dot{\mathbf{f}}_{ij}(\boldsymbol{\beta^0})\dot{\mathbf{f}}_{il}^t(\boldsymbol{\beta^0})| \geq 6I\!\!E[\dot{\mathbf{f}}_{ij}(\boldsymbol{\beta^0})\dot{\mathbf{f}}_{il}^t(\boldsymbol{\beta^0})/\epsilon\big] \leq \epsilon/6$. Using Lemma 1 (iii), we obtain

$$I\!\!P\big[\|\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})\|_1 \geq \frac{6d}{\epsilon}\max_{1\leq j,l\leq d} I\!\!E[\dot{\mathbf{f}}_{ij}(\boldsymbol{\beta^0})\dot{\mathbf{f}}_{il}^t(\boldsymbol{\beta^0})]\big] \leq \frac{\epsilon}{6},$$

relation that involves, since $\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|_1 \leq C\eta$ for $\eta \to 0$, that
$I\!\!P\big[\|\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})\|_1 \geq 6d/\epsilon \max_{1\leq j,l\leq d} I\!\!E[\dot{\mathbf{f}}_{ij}(\boldsymbol{\beta^0})\dot{\mathbf{f}}_{il}^t(\boldsymbol{\beta^0})]\big]$
$\leq I\!\!P\big[\|\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})\|_1 \geq 6d/\epsilon \max_{1\leq j,l\leq d} I\!\!E[\dot{\mathbf{f}}_{ij}(\boldsymbol{\beta^0})\dot{\mathbf{f}}_{il}^t(\boldsymbol{\beta^0})]\big] \leq \epsilon/6$.
Then, for all $\epsilon > 0$

$$I\!\!P\big[\|\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})\|_1 \geq \frac{6d}{\epsilon}\max_{1\leq j,l\leq d} I\!\!E[\dot{\mathbf{f}}_{ij}(\boldsymbol{\beta^0})\dot{\mathbf{f}}_{il}^t(\boldsymbol{\beta^0})]\big] \leq \frac{\epsilon}{6}. \tag{52}$$

For $\ddot{\mathbf{f}}_i(\tilde{\boldsymbol{\beta}}_i)(\boldsymbol{\beta}-\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})$ of relation (46), using assumption (A3) and the Markov's inequality, we obtain for each j-th component $\dot{\mathbf{f}}_{ij}(\boldsymbol{\beta^0})$ of the vector $\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})$, for all $\epsilon_4 > 0$, that $I\!\!P\big[|\dot{\mathbf{f}}_{ij}(\boldsymbol{\beta^0})| \geq \epsilon_4\big] \leq I\!\!E[\dot{\mathbf{f}}_{ij}(\boldsymbol{\beta^0})]/\epsilon_4$. We choose, for all $\epsilon > 0$, $\epsilon_4 = 6I\!\!E[\dot{\mathbf{f}}_{ij}(\boldsymbol{\beta^0})]/\epsilon$ and this last relation becomes $I\!\!P\big[|\dot{\mathbf{f}}_{ij}(\boldsymbol{\beta^0})| \geq 6I\!\!E[\dot{\mathbf{f}}_{ij}(\boldsymbol{\beta^0})]/\epsilon\big] \leq \epsilon/6$. Applying Lemma 1 (i), for all $\epsilon > 0$ we obtain

$$I\!\!P\big[\|\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\|_1 \geq \frac{6d}{\epsilon}\max_{1\leq j\leq d} I\!\!E[\dot{\mathbf{f}}_{ij}(\boldsymbol{\beta^0})]\big] \leq \frac{\epsilon}{6}. \tag{53}$$

Using assumption (A3), and relations (49), (53), we can write that
$I\!\!P\big[\|\ddot{\mathbf{f}}_i(\tilde{\boldsymbol{\beta}}_i)(\boldsymbol{\beta}-\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})\|_1 \geq 6d/\epsilon \max_{1\leq j\leq d} I\!\!E[\dot{\mathbf{f}}_{ij}(\boldsymbol{\beta^0})]\big]$
$\leq I\!\!P\big[\|\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})\|_1 \geq 6d/\epsilon \max_{1\leq j\leq d} I\!\!E[\dot{\mathbf{f}}_{ij}(\boldsymbol{\beta^0})]\big] \leq \epsilon/6$.
Therefore, for all $\epsilon > 0$,

$$I\!\!P\big[\|\ddot{\mathbf{f}}_i(\tilde{\boldsymbol{\beta}}_i)(\boldsymbol{\beta}-\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})\|_1 \geq \frac{6d}{\epsilon}\max_{1\leq j\leq d} I\!\!E[\dot{\mathbf{f}}_{ij}(\boldsymbol{\beta^0})]\big] \leq \frac{\epsilon}{6}. \tag{54}$$

Taking into account assumptions (A3), (A4), by relations (49), (53), we can prove in a similar way as for relation (54) that, for all $\epsilon > 0$,

$$I\!\!P\big[\|\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})^t\ddot{\mathbf{f}}_i(\tilde{\tilde{\boldsymbol{\beta}}}_i)(\boldsymbol{\beta}-\boldsymbol{\beta^0})\|_1 \geq \frac{6d}{\epsilon}\max_{1\leq j\leq d} I\!\!E[\dot{\mathbf{f}}_{ij}(\boldsymbol{\beta^0})]\big] \leq \frac{\epsilon}{6}. \tag{55}$$

For the last term on the right-hand side of (46), since the function $\ddot{\mathbf{f}}(\mathbf{X},\boldsymbol{\beta})$ is bounded, by assumption (A3), we have that, for all $\boldsymbol{\beta} \in \mathcal{V}_\eta(\boldsymbol{\beta}^{\mathbf{0}})$, for all $\epsilon > 0$, there exists $\epsilon_5 > 0$, such that $I\!P[\|\ddot{\mathbf{f}}_i(\tilde{\boldsymbol{\beta}}_i)\|_1\|\ddot{\mathbf{f}}_i(\tilde{\tilde{\boldsymbol{\beta}}}_i)\|_1 \geq \epsilon_5] \leq \epsilon/6$. Using this relation, we show similarly, then, for all $\epsilon > 0$, there exists $\epsilon_5 > 0$, such that,

$$I\!P\big[\|\ddot{\mathbf{f}}_i(\tilde{\boldsymbol{\beta}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})^t\ddot{\mathbf{f}}_i(\tilde{\tilde{\boldsymbol{\beta}}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})\|_1 \geq \epsilon_5\big] \leq \frac{\epsilon}{6}. \tag{56}$$

Choosing

$$M = \sup\left\{\frac{\sigma d}{\sqrt{\epsilon}}\max_{1\leq j\leq d}\sqrt{6V_{jj}}, \epsilon_2, \frac{6d}{\epsilon}\max_{1\leq j,l\leq d}\{I\!E[\dot{\mathbf{f}}_{ij}(\boldsymbol{\beta}^{\mathbf{0}})\dot{\mathbf{f}}^t_{il}(\boldsymbol{\beta}^{\mathbf{0}})], I\!E[\dot{\mathbf{f}}_{ij}(\boldsymbol{\beta}^{\mathbf{0}})]\}, \epsilon_5\right\},$$

and combining (48), (51), (52), (54), (55), (56) together, lemma yields. □

**Lemma 3** *Under the same assumptions of Theorem 2, we have*

$$\frac{1}{n\theta_{nk}}\sum_{i\in I}\mathbf{g}_i(\boldsymbol{\beta}) = O_{I\!P}((n\theta_{nk})^{-1/2}) + \mathbf{V}_1^0(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}) + o_{I\!P}(\boldsymbol{\beta}-\boldsymbol{\beta}^0).$$

*Proof of Lemma 3.* By the Taylor's expansion up to the order 3 of $\mathbf{g}_i(\boldsymbol{\beta})$ at $\boldsymbol{\beta} = \boldsymbol{\beta}^{\mathbf{0}}$, we obtain

$$\begin{aligned}
\frac{1}{n\theta_{nk}}\sum_{i\in I}\mathbf{g}_i(\boldsymbol{\beta}) &= \frac{1}{n\theta_{nk}}\sum_{i\in I}\dot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})\varepsilon_i + \frac{1}{2n\theta_{nk}}\sum_{i\in I}\ddot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})\varepsilon_i \\
&-\frac{1}{2n\theta_{nk}}\sum_{i\in I}\dot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})\dot{\mathbf{f}}^t_i(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}) - \frac{1}{6n\theta_{nk}}\sum_{i\in I}\dot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})^t\ddot{\mathbf{f}}_i(\tilde{\tilde{\boldsymbol{\beta}}}_i)(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}) \\
&-\frac{1}{4n\theta_{nk}}\sum_{i\in I}\ddot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})\dot{\mathbf{f}}^t_i(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}) \qquad (57) \\
&-\frac{1}{12n\theta_{nk}}\sum_{i\in I}\ddot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})^t\ddot{\mathbf{f}}_i(\tilde{\tilde{\boldsymbol{\beta}}}_i)(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}) \\
&+\frac{1}{6n\theta_{nk}}\sum_{i\in I}\mathbf{M}_i\varepsilon_i - \frac{1}{12n\theta_{nk}}\sum_{i\in I}\mathbf{M}_i(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})^t\ddot{\mathbf{f}}_i(\tilde{\tilde{\boldsymbol{\beta}}}_i)(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}),
\end{aligned}$$

where $\mathbf{M}_i = \left(\sum_{l=1}^d\sum_{k=1}^d \frac{\partial^3\dot{\mathbf{f}}_i(\tilde{\boldsymbol{\beta}}_i)}{\partial\boldsymbol{\beta}_j\partial\boldsymbol{\beta}_k\partial\boldsymbol{\beta}_l}(\boldsymbol{\beta}_k-\boldsymbol{\beta}^0_k)(\boldsymbol{\beta}_l-\boldsymbol{\beta}^0_l)\right)_{1\leq j,k,l\leq d}$ is a vector of dimension $(d\times 1)$, and $\tilde{\tilde{\boldsymbol{\beta}}}_i = \boldsymbol{\beta}^{\mathbf{0}} + \mathbf{v}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})$, with $\mathbf{v}\in[0,1]^d$.
For the first term of the right-hand side of (57), by the central limit theorem, and the fact that $I\!E[\mathbf{g}_i(\boldsymbol{\beta}^{\mathbf{0}})] = 0$, we have

$$(n\theta_{nk})^{-1}\sum_{i\in I}\mathbf{g}_i(\boldsymbol{\beta}^{\mathbf{0}}) = O_{I\!P}((n\theta_{nk})^{-1/2}). \tag{58}$$

For the second term of the right-hand side of (57), by the law of large numbers, the term $(n\theta_{nk})^{-1}\sum_{i\in I}\ddot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})\varepsilon_i$ converges almost surely to the expected of $\ddot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})\varepsilon_i$ as $n \to \infty$. Furthermore, since $\varepsilon_i$ is independent of $\mathbf{X}_i$ and $I\!E[\varepsilon_i]=0$, we have

$$\frac{1}{n\theta_{nk}} \sum_{i \in I} \ddot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})\varepsilon_i = o_{I\!P}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}). \tag{59}$$

For the third term of the right-hand side of (57), by the law of large numbers and assumption (A4), the term $(n\theta_{nk})^{-1}\sum_{i\in I}\dot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})$ converges almost surely to the expected value of $\dot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})$ as $n\to\infty$. On the other hand, since $(n\theta_{nk})^{-1}\sum_{i\in I}\ddot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})\varepsilon_i \overset{a.s}{\longrightarrow} 0$, we have

$$\frac{1}{n\theta_{nk}} \sum_{i \in I} \dot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}) = -\mathbf{V}_1^0(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})(1+o_{I\!P}(1)). \tag{60}$$

For the fourth term of the right-hand side of (57), by the law of large numbers, using assumption (A3) and the relation (53), we can write $(6n\theta_{nk})^{-1}\|\sum_{i\in I}\dot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})$ $(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})^t\ddot{\mathbf{f}}_i(\tilde{\tilde{\boldsymbol{\beta}}}_i)(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})\|_1 = O_{I\!P}(\|\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}\|_2^2)$, which implies

$$\frac{1}{6n\theta_{nk}} \sum_{i \in I} \dot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})^t\ddot{\mathbf{f}}_i(\tilde{\tilde{\boldsymbol{\beta}}}_i)(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}) = o_{I\!P}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}). \tag{61}$$

In the same way, using assumption (A3) and relation (53), we obtain, for the fifth term on the right-hand side of (57), that

$$\frac{1}{4n\theta_{nk}} \sum_{i \in I} \ddot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}) = o_{I\!P}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}). \tag{62}$$

For the sixth term of the right-hand side of (57), using the assumption (A3), we have

$$\frac{1}{12n\theta_{nk}} \sum_{i \in I} \ddot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})^t\ddot{\mathbf{f}}_i(\tilde{\tilde{\boldsymbol{\beta}}}_i)(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}) = o_{I\!P}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}). \tag{63}$$

For $1\le j\le d$, and for any fixed $i$, such that $1\le i\le n\theta_{nk}$ , denote by $M_{ij}$ the following random variable designates the j-th component of the vector $\mathbf{M}_i$, such that

$$\mathbf{M}_{ij} = \sum_{l=1}^{d}\sum_{k=1}^{d} \frac{\partial^3 f_i(\tilde{\boldsymbol{\beta}}_i)}{\partial\boldsymbol{\beta}_k\partial\boldsymbol{\beta}_l\partial\boldsymbol{\beta}_j}(\boldsymbol{\beta}_k-\boldsymbol{\beta}_k^0)(\boldsymbol{\beta}_l-\boldsymbol{\beta}_l^0).$$

From assumption (A3), since the function $f^{(3)}(\mathbf{x},\boldsymbol{\beta})$ is bounded for all $\boldsymbol{\beta}\in\mathcal{V}_\eta(\boldsymbol{\beta}^{\mathbf{0}})$, we have with a probability one, $|\mathbf{M}_{ij}|\le C\|\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}\|_2^2$. Applying Lemma 1 (i), we obtain

$$\|\mathbf{M}\|_1 \le C\|\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}\|_2^2. \tag{64}$$

For the term $(6n\theta_{nk})^{-1}\sum_{i\in I}\mathbf{M}_i\varepsilon_i$, using relations (50) and (64), we have $(6n\theta_{nk})^{-1}\|\sum_{i\in I}\mathbf{M}_i\varepsilon_i\|_1 \leq (6n\theta_{nk})^{-1}\sum_{i\in I}\|\mathbf{M}_i\|_1|\varepsilon_i| \leq C(6n\theta_{nk})^{-1}n\theta_{nk}\|\boldsymbol{\beta}-\boldsymbol{\beta^0}\|_2^2 = C\|\boldsymbol{\beta}-\boldsymbol{\beta^0}\|_2^2$. Then,

$$\frac{1}{6n\theta_{nk}}\sum_{i\in I}\mathbf{M}_i\varepsilon_i = o_{I\!P}(\boldsymbol{\beta}-\boldsymbol{\beta^0}). \tag{65}$$

Finally, for the last term of the right-hand side of (57), using assumption (A3) and relation (64), we obtain with probability 1, $(12n\theta_{nk})^{-1}\|\sum_{i\in I}\mathbf{M}_i(\boldsymbol{\beta}-\boldsymbol{\beta^0})^t\ddot{\mathbf{f}}_i(\tilde{\tilde{\boldsymbol{\beta}}}_i)(\boldsymbol{\beta}-\boldsymbol{\beta^0})\|_1 \leq C\|\boldsymbol{\beta}-\boldsymbol{\beta^0}\|_2^2$, which gives,

$$\frac{1}{12n\theta_{nk}}\sum_{i\in I}\mathbf{M}_i(\boldsymbol{\beta}-\boldsymbol{\beta^0})^t\ddot{\mathbf{f}}_i(\tilde{\tilde{\boldsymbol{\beta}}}_i)(\boldsymbol{\beta}-\boldsymbol{\beta^0}) = o_{I\!P}(\boldsymbol{\beta}-\boldsymbol{\beta^0}). \tag{66}$$

Then, combining relations (58), (59), (60), (61), (62), (63), (65) and (66), we obtain lemma. □

**Lemma 4** *Under the same assumptions as in Theorem 3, for all $\varrho > 0$, there exist two positive constants $C = C(\varrho)$, $T = T(\varrho)$ such that*
$I\!P[\max_{\frac{T}{n}\leq\theta_{nk}\leq 1-\frac{T}{n}}(n\theta_{nk}/\log\log n\theta_{nk})^{1/2}\|\frac{\hat{\boldsymbol{\lambda}}(\theta_{nk})}{\min\{\theta_{nk},1-\theta_{nk}\}}\|_2 > C] \leq \varrho$,
$I\!P[\max_{\frac{T}{n}\leq\theta_{nk}\leq 1-\frac{T}{n}}(n\theta_{nk}/\log\log n\theta_{nk})^{1/2}\|\hat{\boldsymbol{\beta}}(\theta_{nk})-\boldsymbol{\beta^0}\|_2 > C] \leq \varrho$,
$I\!P[n^{-1/2}\max_{\frac{T}{n}\leq\theta_{nk}\leq 1-\frac{T}{n}} n\theta_{nk}\|\frac{\hat{\boldsymbol{\lambda}}(\theta_{nk})}{\min\{\theta_{nk},1-\theta_{nk}\}}\|_2 > C] \leq \varrho$,
$I\!P[n^{-1/2}\max_{\frac{T}{n}\leq\theta_{nk}\leq 1-\frac{T}{n}} n\theta_{nk}\|\hat{\boldsymbol{\beta}}(\theta_{nk})-\boldsymbol{\beta^0}\|_2 > C] \leq \varrho$.

*Proof of Lemma 4.* The proof of this lemma is similar to that of Lemma 1.2.2 of [8]. □

In order, to prove Lemma 5, we consider $R_k = n\sigma^{-2}\theta_{nk}(1-\theta_{nk})(\overline{\mathbf{w}}_1-\overline{\mathbf{w}}_2)^t\mathbf{V}^{-1}(\overline{\mathbf{w}}_1-\overline{\mathbf{w}}_2)$. Recall that $\mathbf{V} \equiv I\!E[\dot{\mathbf{f}}(\mathbf{X}_i,\boldsymbol{\beta^0})\dot{\mathbf{f}}^t(\mathbf{X}_i,\boldsymbol{\beta^0})]$, for all $i = 1, ..., n$.

**Lemma 5** *Suppose that the assumptions (A1)-(A4) hold. Under the null hypothesis $H_0$, for all $0 \leq \alpha < 1/2$ we have*
*(i)* $n^\alpha \max_{\theta_{nk}\in\Theta_{nk}}[\theta_{nk}(1-\theta_{nk})]^\alpha|Z_{nk}(\theta_{nk},\hat{\boldsymbol{\lambda}}(\theta_{nk}),\hat{\boldsymbol{\beta}}(\theta_{nk}))-R_k| = O_{I\!P}(1)$.
*(ii)* $\max_{\theta_{nk}\in\Theta_{nk}}[\theta_{nk}(1-\theta_{nk})]|Z_{nk}(\theta_{nk},\hat{\boldsymbol{\lambda}}(\theta_{nk}),\hat{\boldsymbol{\beta}}(\theta_{nk}))-R_k| = O_{I\!P}(n^{-1/2}(\log\log n)^{3/2})$.

*Proof of Lemma 5.* For the score function $\boldsymbol{\phi}_{1n}$ of relation (12), the two terms of the right-hand side are replaced by their decomposition obtained by the relations (21) and (24). On the other hand, we have $\boldsymbol{\phi}_{1n}(\theta_{nk},\hat{\boldsymbol{\lambda}}(\theta_{nk}),\hat{\boldsymbol{\beta}}(\theta_{nk})) = \mathbf{0}_d$. Then, we can write $[\frac{1}{n\theta_{nk}}\sum_{i\in I}\mathbf{g}_i(\boldsymbol{\beta^0})+\mathbf{V}_1^0(\hat{\boldsymbol{\beta}}(\theta_{nk})-\boldsymbol{\beta^0})-\frac{1}{n\theta_{nk}^2}\sum_{i\in I}\mathbf{g}_i(\boldsymbol{\beta^0})\mathbf{g}_i^t(\boldsymbol{\beta^0})$ $\cdot\hat{\boldsymbol{\lambda}}(\theta_{nk})](1+o_{I\!P}(1))-\mathbf{V}_1^0(\mathbf{V}_2^0)^{-1}[\frac{1}{n(1-\theta_{nk})^2}\mathbf{V}_1(\mathbf{V}_2^0)^{-1}\sum_{j\in J}\mathbf{g}_j(\boldsymbol{\beta^0})\mathbf{g}_j^t(\boldsymbol{\beta^0})\hat{\boldsymbol{\lambda}}(\theta_{nk})+$

$\frac{1}{n(1-\theta_{nk})} \cdot \sum_{j \in J} \mathbf{g}_j(\boldsymbol{\beta}^{\mathbf{0}}) + \mathbf{V}_2^0(\hat{\boldsymbol{\beta}}(\theta_{nk}) - \boldsymbol{\beta}^{\mathbf{0}})](1 + o_{I\!P}(1)) = \mathbf{0}_d$. Hence,

$$\begin{aligned}\hat{\boldsymbol{\lambda}}(\theta_{nk}) = &\Big(\frac{1}{\theta_{nk}}\mathbf{D}_{n1}^0 + \frac{1}{1-\theta_{nk}}(\mathbf{V}_1^0(\mathbf{V}_2^0)^{-1})(\mathbf{V}_1^0(\mathbf{V}_2^0)^{-1})^t\mathbf{D}_{n2}^0\Big)^{-1} \\ &\cdot\Big(\frac{1}{n\theta_{nk}}\sum_{i\in I}\mathbf{g}_i(\boldsymbol{\beta}^{\mathbf{0}}) - \frac{\mathbf{V}_1^0(\mathbf{V}_2^0)^{-1}}{n(1-\theta_{nk})}\sum_{j\in J}\mathbf{g}_j(\boldsymbol{\beta}^{\mathbf{0}})\Big)(1+o_{I\!P}(1)) \\ &+ o_{I\!P}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}),\end{aligned}$$

with the matrices $\mathbf{D}_1^0$ and $\mathbf{D}_2^0$ given by relation (23).
On the other hand, by the law of large numbers, we have $-\mathbf{V}_1^0 \overset{a.s}{\longrightarrow} \mathbf{V}$ and $-\mathbf{V}_2^0 \overset{a.s}{\longrightarrow} \mathbf{V}$. Then, $\mathbf{V}_1^0(\mathbf{V}_2^0)^{-1} \overset{a.s}{\longrightarrow} I_d$. Always, by the law of large numbers, $\mathbf{D}_{n1}^0$ and $\mathbf{D}_{n2}^0$ converge almost surely to $\sigma^2\mathbf{V}$ as $n \to \infty$.
By Theorem 2, we proved that $\hat{\boldsymbol{\lambda}}(\theta_{nk}) = \theta_{nk}O_p((n\theta_{nk})^{-1/2})$. Then, we obtain

$$\hat{\boldsymbol{\lambda}}(\theta_{nk}) = \sigma^{-2}\theta_{nk}(1-\theta_{nk})\mathbf{V}^{-1}(\overline{\mathbf{w}}_1 - \overline{\mathbf{w}}_2)(1+o_{I\!P}(1)) + o_{I\!P}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}). \quad (67)$$

The limited development of the statistic $Z_{nk}(\theta_{nk}, \hat{\boldsymbol{\lambda}}(\theta_{nk}), \hat{\boldsymbol{\beta}}(\theta_{nk}))$, specified by the relation (11), in the neighbourhood of $(\boldsymbol{\lambda}, \boldsymbol{\beta}) = (0, \boldsymbol{\beta}^{\mathbf{0}})$ up to order 2, can be written
$\Big[\frac{2\hat{\boldsymbol{\lambda}}^t(\theta_{nk})}{\theta_{nk}}\sum_{i\in I}\mathbf{g}_i(\boldsymbol{\beta}^{\mathbf{0}}) - \frac{2\hat{\boldsymbol{\lambda}}^t(\theta_{nk})}{1-\theta_{nk}}\mathbf{V}_1^0(\mathbf{V}_2^0)^{-1}\sum_{j\in J}\mathbf{g}_i(\boldsymbol{\beta}^{\mathbf{0}})\Big] - \Big[\frac{\hat{\boldsymbol{\lambda}}^t(\theta_{nk})}{(1-\theta_{nk})^2}\mathbf{V}_1^0(\mathbf{V}_2^0)^{-1}$ $\cdot \sum_{j\in J}\mathbf{g}_j(\boldsymbol{\beta}^{\mathbf{0}})\mathbf{g}_j^t(\boldsymbol{\beta}^{\mathbf{0}})\mathbf{V}_1^0(\mathbf{V}_2^0)^{-1}\hat{\boldsymbol{\lambda}}(\theta_{nk}) + \frac{\hat{\boldsymbol{\lambda}}^t(\theta_{nk})}{\theta_{nk}^2}\sum_{i\in I}\mathbf{g}_i(\boldsymbol{\beta}^{\mathbf{0}})\mathbf{g}_i^t(\boldsymbol{\beta}^{\mathbf{0}})\hat{\boldsymbol{\lambda}}(\theta_{nk})\Big] +$ $\Big[2\hat{\boldsymbol{\lambda}}^t(\theta_{nk}) \cdot (\hat{\boldsymbol{\beta}}(\theta_{nk}) - \boldsymbol{\beta}^{\mathbf{0}})\Big(-\frac{1}{1-\theta_{nk}}\mathbf{V}_1^0(\mathbf{V}_2^0)^{-1}\sum_{j\in J}\dot{\mathbf{g}}_j(\boldsymbol{\beta}^{\mathbf{0}})\mathbf{V}_1^0(\mathbf{V}_2^0)^{-1} + \frac{1}{\theta_{nk}}$ $\cdot \sum_{i\in I}\dot{\mathbf{g}}_i(\boldsymbol{\beta}^{\mathbf{0}}) + \frac{\partial[\mathbf{V}_1^0(\mathbf{V}_2^0)^{-1}]}{\partial\boldsymbol{\beta}}\sum_{j\in J}\mathbf{g}_j(\boldsymbol{\beta}^{\mathbf{0}})\Big)\Big]$.
Replacing $\hat{\boldsymbol{\lambda}}(\theta_{nk})$ in the first term on the right-hand side of the last relation, by the value obtained in (67) ,we find that this term is equal to $2n\sigma^{-2}\theta_{nk}(1-\theta_{nk})(\overline{\mathbf{w}}_1-\overline{\mathbf{w}}_2)^t\mathbf{V}^{-1}(\overline{\mathbf{w}}_1-\overline{\mathbf{w}}_2) + o_{I\!P}(\|\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}\|_2)$. Similarly, using the fact that $\mathbf{D}_{n1}^0$ and $\mathbf{D}_{n2}^0$ converge to $\sigma^2\mathbf{V}$, as $n \to \infty$, we can demonstrate that the second term is equal to $n\sigma^{-2}\theta_{nk}(1-\theta_{nk})(\overline{\mathbf{w}}_1-\overline{\mathbf{w}}_2)^t\mathbf{V}^{-1}(\overline{\mathbf{w}}_1-\overline{\mathbf{w}}_2) + o_{I\!P}(\|\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}\|_2)$. Finally, by the central limit theorem, we have that $(n(1-\theta_{nk}))^{-1}\sum_{j\in J}\mathbf{g}_j(\boldsymbol{\beta}^0) = O_{I\!P}((n\theta_{nk})^{-1/2})$. We obtain that the third term is $o_{I\!P}(n\sigma^{-2}\theta_{nk}(1-\theta_{nk})(\overline{\mathbf{w}}_1-\overline{\mathbf{w}}_2)^t\mathbf{V}^{-1}(\overline{\mathbf{w}}_1-\overline{\mathbf{w}}_2))$. Combining the obtained results, we have $Z(\theta_{nk}, \hat{\boldsymbol{\lambda}}(\theta_{nk}), \hat{\boldsymbol{\beta}}(\theta_{nk})) = \Big(n\sigma^{-2}\theta_{nk}(1-\theta_{nk})(\overline{\mathbf{w}}_1-\overline{\mathbf{w}}_2)^t\mathbf{V}^{-1}(\overline{\mathbf{w}}_1-\overline{\mathbf{w}}_2)\Big)\Big(1+o_{I\!P}(1)\Big) + O_{I\!P}(\|\hat{\boldsymbol{\beta}}(\theta_{nk})-\boldsymbol{\beta}^{\mathbf{0}}\| + \|\hat{\boldsymbol{\lambda}}(\theta_{nk})/\min\{\theta_{nk}, 1-\theta_{nk}\}\|)$. This last relation, together with Lemma 4 imply Lemma 5. □

## References


1. Bai, J. Likelihood ratio tests for multiple structural changes. *Journal of Econometrics*, **91**, 299-323, 1999.
2. Bai, J., Perron, P. Estimating and testing linear models with multiple structural changes. *Econometrica*, **1**, 47-48, 1998.

3. Boldea, O., Hall, A.R. Estimation and inference in unstable nonlinear least squares models. *Journal of Econometrics*, **172**(1), 158-167, 2013.
4. Ciuperca, G. A general criterion to determinate the number of change-points. *Statistics and Probability Letters*, **81**(8), 1267-1275, 2011.
5. Ciuperca, G. Penalized least absolute deviations estimation for nonlinear model with change-points. *Statistical Papers*, **52**(2), 371-390, 2011.
6. Ciuperca G. Two tests for sequential detection of a change-point in a nonlinear model, *Journal of Statistical Planning and Inference*, **143**(10), 1621-1834, 2013.
7. Ciuperca G. Empirical likelihood for nonlinear model with missing responses, *Journal of Statistical Computation and Simulation*, **83**(4), 737-756, 2013.
8. Csörgö M., Horváth L. Limit Theorems in Change-point Analysis, John Wiley and Sons (1997).
9. Hall A., Sen A. Structural stability testing in models estimated by generalized method of moments, *Journal of Business Economic Statistics*, **17**(3), 335-348, 1999.
10. Horváth, L., Hušková, M., Kokoszka, P., Steinebach, J. Monitoring changes in linear models. *Journal of Statistical Planning and Inference*, **126**(1), 225-251, 2004.
11. Hušková, M., Kirch, C. Bootstrapping sequential change-point tests for linear regression. *Metrika 75*, no 5, 673-708, 2012.
12. Lai, T. L., Xing, H. Sequential change-point detection when the pre- and post-change parameters are unknown. *Sequential Analysis,* **29**(2), 162–175, 2010.
13. Lee, S., Seo, M.H., Shin, Y. Testing for threshold effects in regression models. *Journal of American Statistical Association,***493**, 220–231, 2011.
14. Liu, Y., Zou, C., Zhang, R. Empirical likelihood ratio test for a change-point in linear regression model. *Communications in Statistics-Theory and Methods*, **37**, 2551-2563, 2008.
15. Mei, Y. Sequential change-point detection when unknown parameters are present in the pre-change distribution. *Annals of Statistics* **34**(1), 92-122, 2006.
16. Neumeyer, N., Van Keilegom, I. Change-point tests for the error distribution in non-parametric regression. *Scandinavian Journal of Statistics,* **36**(3), 518-541, 2009.
17. Nosek, K. Schwarz information criterion based tests for a change-point in regression models. *Statistical Papers*, **51**(4), 915-929, 2010.
18. Qu, Z. Testing for structural change in regression quantiles *Journal of Econometrics*, **146**, 170-184, 2008.
19. Seber G., Wild C. Testing for structural change in regression quantiles, Wiley series in probability and mathematical Statistics, Wiley, Hoboken, NJ, 2003.
20. Van der Vaart A., Wellner J. Weak Convergence and Empirical processes, 1996.
21. Wu Y. Simultaneous change point analysis and variable selection in a regression problem. *Journal of Multivariate Analysis*, **99**, 2154-2171, 2008.
22. Yu, W., Niu, C., Xu, W. An empirical likelihood inference for the coefficient difference of a two-sample linear model with missing response data. *Metrika*, DOI 10.1007/s00184-013-0459-3.
23. Zi, X., Zou, C., Liu, Y. Two-sample empirical likelihood method for difference between coefficients in linear regression model. *Statistical Papers*, **53**(1), 83-93, 2012.
24. Zou C., Liu Y., Qin P., Wang Z. Empirical likelihood ratio test for a change point. *Statistics and Probability Letters*, **77**, 374-382, 2007.